МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Физико-технический факультет**

**Кафедра теоретической физики и компьютерных технологий**

**КУРСОВОЙ ПРОЕКТ**

**Анализ возможностей технологий машинного обучения**

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Бирюков Дмитрий Михайлович

Курс 2

Направление 09.03.02 Информационные системы и технологии

Научный руководитель

преподаватель\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.Н.Занчко

Нормоконтролер инженер\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Г.Д.Цой

Краснодар 2018

|  |  |
| --- | --- |
| Введение. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 3 |
| 1 История машинного обучения. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 6 |
| 2 Основные методы и возможности машинного обучения . . . . . . . . . . . . . | 9 |
| 2.1 Основные методы машинного обучения. . . . . . . . . . . . . . . . . . . . . . . | 9 |
| 2.1.1 Контролируемое обучение. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 9 |
| 2.1.1.1 Наивная байесовская классификация. . . . . . . . . . . . . . . . . | 9 |
| 2.1.1.2 Метод наименьших квадратов. . . . . . . . . . . . . . . . . . . . . . . | 10 |
| 2.1.1.3 Логистическая регрессия. . . . . . . . . . . . . . . . . . . . . . . . . . . | 10 |
| 2.1.1.4 Метод опорных векторов. . . . . . . . . . . . . . . . . . . . . . . . . . . | 12 |
| 2.1.1.5 Метод ансамблей. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 12 |
| 2.1.2 Неконтролируемое обучение. . . . . . . . . . . . . . . . . . . . . . . . . . . . | 14 |
| 2.1.2.1 Алгоритмы кластеризации.. . . . . . . . . . . . . . . . . . . . . . . . . | 14 |
| 2.1.2.2 Метод главных компонент.. . . . . . . . . . . . . . . . . . . . . . . . . | 15 |
| 2.1.2.3 Сингулярное разложение.. . . . . . . . . . . . . . . . . . . . . . . . . . | 16 |
| 2.1.2.4 Анализ независимых компонент. . . . . . . . . . . . . . . . . . . . . | 16 |
| 2.2 Возможности машинного обучения. . . . . . . . . . . . . . . . . . . . . . . . . . | 17 |
| 2.2.1 Кредитный скоринг. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 17 |
| 2.2.2 Принятие решений. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 17 |
| 2.2.3 Извлечение информации. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 18 |
| 2.2.4 Защита от мошенничества. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 18 |
| 2.2.5 Алгоритмические стратегии для торговли. . . . . . . . . . . . . . . . . | 19 |
| Заключение. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 21 |
| Список использованных источников. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 23 |

**ВВЕДЕНИЕ**

Машинное обучение - этот термин, скорее всего, встречался вам много раз, его часто используют как синоним искусственного интеллекта, но на самом деле машинное обучение – это лишь верхушка айсберга. Согласно книге [1], в настоящее время мы сталкиваемся с машинным обучением каждый день, даже не подозревая об этом. Когда вы, например, просматриваете свой почтовый ящик, наибольший поток спама проходит мимо вас, за счет того что он был отфильтрован с помощью механизмов машинного обучения. Благодаря машинному обучению поисковая система распознает, какую рекламу показывать в ответ на ваш запрос в поисковике. Голосовые помощники Baidu и Google, рекомендации в Яндекс, распознавание лиц в Facebook и iPhone X . В последние десятилетие машинное обучение получило резкий толчок благодаря развитию науки, увеличению вычислительных мощностей. Обычно, когда машина выполняет какую-то задачу, все ее действия выполняются по определенному программистом-разработчиком алгоритму. Машина, использующая алгоритм машинного обучения, угадывает все сама, делает вывод на основе входных данных, и чем больше этих данных, тем точнее результат ее работы. Таким образом, машина программирует сама себя.

Актуальность исследования поставленной проблемы обусловлена автоматизацией решений сложных профессиональных задач в самых разных областях человеческой деятельности таких как:

* Диагностика в медицине
  + Кредитный скоринг
  + Предсказание ухода клиентов
  + Обнаружение мошенничества
  + Биржевой технический анализ
  + Биржевой надзор
  + Техническая диагностика
  + Робототехника
  + Компьютерное зрение
  + Распознавание речи
  + Распознавание текста
  + Обнаружение спама
  + Категоризация документов
  + Распознавание рукописного ввода

Сфера, где применяется машинное обучение, постоянно увеличивается. Информатизация общества приводит к накоплению больших объёмов данных в производстве, науке, транспорте, бизнесе, медицине.

Объектом исследования являются возможности технологий машинного обучения.

Предметом – различные технологии машинного обучения.

Целью исследования является анализ возможностей технологий машинного обучения. Достижение поставленной цели предусматривает постановку следующих задач:

- изучение истории машинного обучения;

- определение основных способов машинного обучения;

- исследование возможностей машинного обучения;

- рассмотрение практической сферы применения.

Теоретико-методологическую основу исследования составляют труды Педро Домингоса.

Эмпирическую базу исследования составили: книги по Machine Learning,различные статьи в Интернет.

Структура курсовой работы обусловлена целью и задачами исследования и включает в себя: введение, два раздела, заключение, список литературы.

В первом разделе рассматривается история машинного обучения.

Во втором разделе анализируется основные способы и возможности машинного обучения.

**1 История машинного обучения**

Как и многое другое в искусственном интеллекте, история ML началась с казалось бы многообещающих работ в 1950-х — 1960-х годах, а затем последовал длительный период накопления знаний, известный как «зима искусственного интеллекта». В самые последние годы наблюдается взрывной интерес главным образом к одному из направлений — глубинному, или глубокому обучению (deep leаrning).

Первопроходцами ML были Артур Сэмюэль, Джозеф Вейцбаум и Фрэнк Розенблатт. Первый получил широкую известность созданием в 1952 году самообучающейся программы Checkers-playing, умевшей, как следует из названия, играть в шашки. Возможно, более значимым для потомков оказалось его участие вместе с Дональдом Кнутом в проекте TeX, результатом которого стала система компьютерной верстки, вот уже почти 40 лет не имеющая себе равных для подготовки математических текстов. Второй в 1966 году написал виртуального собеседника ELIZA, способного имитировать (а скорее, пародировать) диалог с психотерапевтом; очевидно, что своим названием программа обязана героине из пьесы Бернарда Шоу. А дальше всех пошел Розенблатт, он в конце 50-х в Корнелльском университете построил систему Mark I Perceptron, которую можно признать первым нейрокомпьютером.

Mark I предназначался для классификации визуальных образов (символов алфавита) и представлял собой электронно-механическую систему, центром которой были 400 управляемых фото-сенсоров, они-то и служили моделью сетчатки. Управлялись они с помощью манипуляторов на шаговых моторах. На распространенной в Сети фотографии рядом с Розенблаттом видна небольшая коробка, которую обычно и называют Mark I, хотя это лишь один из узлов, в полном же комплекте система занимала шесть солидных стоек, она сохранилась в Смитсоновском музее в Вашингтоне. В дополнение к ней для распознавания речи была предпринята попытка создания еще более громоздкой системы Tobermory Perceptron. Обе системы были практическими реализациями пусть простых, но все же нейронных сетей. Сложно сказать, как бы продолжилось это направление развития, не будь оно прервано двумя обстоятельствами — смертью Розенблатта, он погиб на яхте, и ожесточенной критикой со стороны Марвина Минского и Сеймура Паперта, доказывавших, что перцептрон неспособен к обучению.

За последующие 30-40 лет усилиями академически ориентированных ученых машинное обучение было превращено в самостоятельную математическую дисциплину.

Начало первого десятилетия XXI века оказалось поворотной точкой в истории ML, и объясняется это тремя синхронными тенденциями, давшими в совокупности заметный синергетический эффект.

Первая — Большие Данные. Данных стало так много, что новые подходы были вызваны к жизни не любознательностью ученых, а практической необходимостью.

Вторая — снижение стоимости параллельных вычислений и памяти. Эта тенденция обнаружилась в 2004 году, когда компания Google раскрыла свою технологию MapReduce, за которой последовал ее открытый аналог Hadoop (2006), и совместно они дали возможность распределить обработку огромных объемов данных между простыми процессорами. Тогда же Nvidia совершила прорыв на рынке GPU: если раньше в игровом сегменте ей могла составить конкуренцию AMD/ATI, то в сегменте графических процессоров, которые можно использовать для целей машинного обучения, она оказалась монополистом. И в то же время заметно уменьшилась стоимость оперативной памяти, что открыло возможность для работы с большими объемами данных в памяти и, как следствие, появились многочисленные новые типы баз данных, в том числе NoSQL. И, наконец, в 2014 году появился программный каркас Apache Spark для распределенной обработки неструктурированных и слабоструктурированных данных, он оказался удобен для реализации алгоритмов машинного обучения.

Третья — новые алгоритмы глубинного машинного обучения, наследующие и развивающие идею перцептрона в сочетании с удачной научной PR-кампанией. Своей критикой Марвин Минский и Сеймур Паперт сыграли положительную роль, они вскрыли слабости перцептрона в том виде, как его придумал Розенблатт, и одновременно стимулировали дальнейшие работы по нейронным сетям, до 2006 года остававшиеся теоретическими. Вероятно, первым, кто решил «развинтить» (углубить) перцепрон, был советский математик А.Г. Ивахненко, опубликовавший начиная с 1965 года ряд статей и книг, в которых, в частности, описана моделирующая система «Альфа». В 1980 году Кунихика Фукусима предложил иерархическую многослойную сверточную нейронную сеть, известную как неокогнитрон.

Последующие годы отмечены интенсивной работой многих ученых в области глубинных нейронных сетей (Deep Neural Network, DNN), однако детальный и желательно объективный анализ событий этого периода еще ждет своего исследователя.

**2 Основные методы и возможности машинного обучения**

**2.1 Основные методы машинного обучения**

Методы машинного обучения можно разделить на 3 основные категории: контролируемое, неконтролируемое и подкрепляемое обучение. Контролируемое обучение полезно в тех случаях, когда свойство (ярлык) доступно для определенного массива данных (обучающего набора), но на данный момент оно отсутствует и должно быть предсказано для других случаев. Неконтролируемое обучение используется для обнаружения неявных отношений в данном немаркированном наборе данных. Подкрепляемое обучение - что-то среднее между вышеописанными категориями: есть некоторая форма обратной связи, доступная для каждого шага или действия, но отсутствует ярлык и сообщение об ошибке.

**2.1.1 Контролируемое обучение**

**2.1.1.1 Наивная байесовская классификация**

Наивные байесовские классификаторы представляют собой семейство простых вероятностных классификаторов, которые основаны на применении Теоремы Байеса со строгими (наивными) предположениями о независимости функций. Из книги [2] P (A\B) является вероятностью гипотезы A при наступлении события B (апостериорная вероятность), P (B\A) — вероятностью наступления события B при истинности гипотезы A, P (A) — априорной вероятностью гипотезы A и P (B) — полной вероятностью наступления события B.

Абстрагируясь от теории и переходя к практике, можно выделить следующие сферы применения Теоремы Байеса:

* «отлов» спама в электронной почте;
* сегментация новостных статей по их тематике;
* определение эмоционального окраса блока текста;
* программное обеспечение для распознавания лиц.

### 2.1.1.2 Метод наименьших квадратов

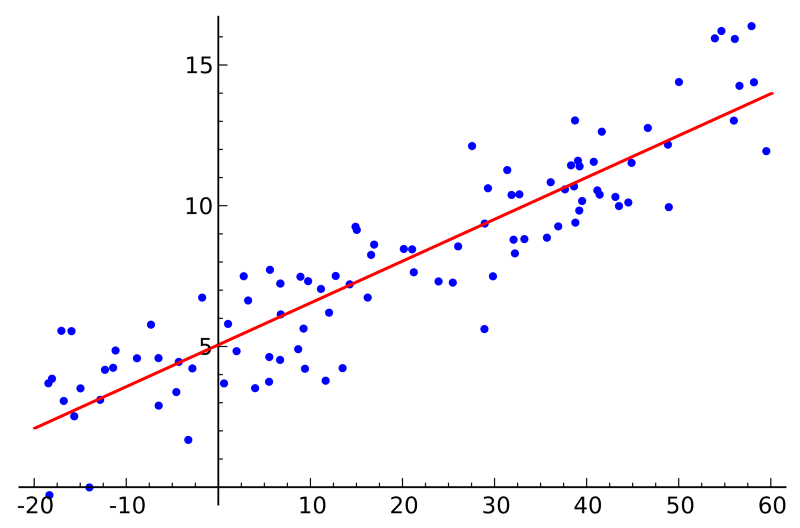


Рисунок 2

Если вы знакомы со статистикой, то наверняка слышали о линейной регрессии ранее. Наименьшие квадраты выступают в роли метода для реализации линейной регрессии. Чаще всего она представляется в виде задачи подгонки прямой линии, проходящей через множество точек. Есть несколько вариантов ее осуществления, и метод наименьших квадратов — один из них. Можно нарисовать линию, а затем измерить расстояние по вертикали от каждой точки к линии и «перенести» эту сумму вверх. Необходимой линией будет та конструкция, где сумма расстояний будет минимальной. Иными словами, кривая проводится через точки, имеющие нормально распределенное отклонение от истинного значения(рис.2).

Если линейная функция применима для подбора данных, то метод наименьших квадратов относится к типам метрики ошибок, которая минимизирует погрешности.

### 2.1.1.3 Логистическая регрессия

Логистическая регрессия представляет собой мощный статистический способ прогнозирования вероятности возникновения некоторого события с одной или несколькими независимыми переменными(рис.3). Логистическая регрессия определяет степень зависимости между категориальной зависимой и одной или несколькими независимыми переменными путем использования логистической функции, являющейся аккумулятивным логистическим распределением.

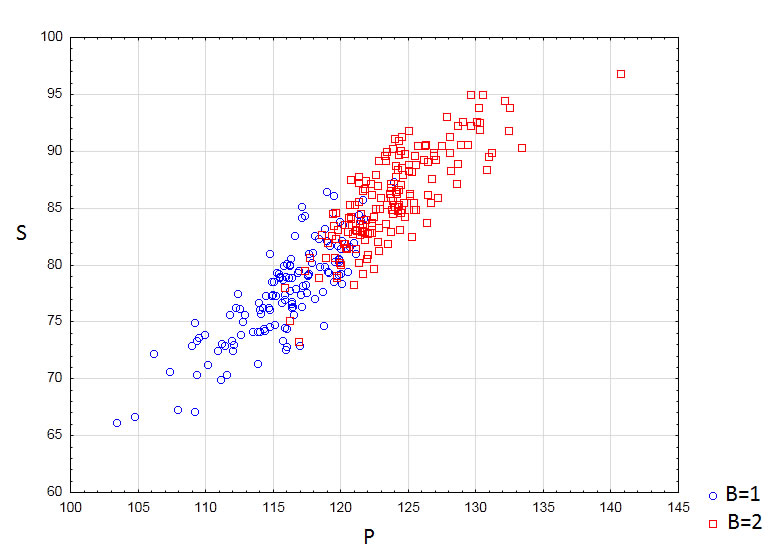


Рисунок 3

Данный алгоритм активно используется в реальной жизни, а именно при:

* оценке кредитоспособности лица (кредитном скоринге);
* измерении показателей успешности маркетинговых кампаний;
* предсказании доходов с определенного продукта;
* вычислении возможности возникновения землетрясения в конкретный день.

### 2.1.1.4 Метод опорных векторов

Метод опорных векторов (SVM) — это набор алгоритмов, использующихся для задач классификации и регрессионного анализа(рис.4). Учитывая, что в N-мерном пространстве каждый объект принадлежит одному из двух классов, SVM генерирует (N-1)-мерную гиперплоскость с целью разделения этих точек на 2 группы. Это как если бы вы на бумаге изобразили точки двух разных типов, которые можно линейно разделить. Помимо того, что метод выполняет сепарацию объектов, SVM подбирает гиперплоскость так, чтобы та характеризовалась максимальным удалением от ближайшего элемента каждой из групп.

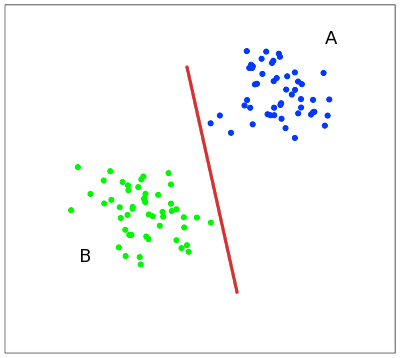


Рисунок 4

Среди наиболее масштабных проблем, которые были решены с помощью метода опорных объектов (и его модифицированных реализаций) выделяют отображение рекламных баннеров на сайтах, распознавание пола на основании фотографии и сплайсинг человеческой ДНК.

### 2.1.1.5 Метод ансамблей

Метод ансамблей, следуя из содержания книги [3], основан на обучающих алгоритмах, которые формируют множество классификаторов, а затем сегментируют новые точки данных, отталкиваясь от голосования или усреднения. Оригинальный метод ансамблей — не что иное, как Байесовское усреднение, но более поздние алгоритмы включают исправления ошибок выходного кодирования, бэггинг (bagging) и бустинг (boosting). Бустинг направлен на превращение слабых моделей в сильные путем построения ансамбля классификаторов. Бэггинг также агрегирует усовершенствованные классификаторы, но используется при этом параллельное обучение базовых классификаторов. Говоря языком математической логики, бэггинг — улучшающее объединение, а бустинг — улучшающее пересечение.

Все же, почему метод ансамблей превосходит отдельно стоящие прогнозные модели?

* Он минимизирует влияние случайностей. Агрегированный классификатор «усредняет» ошибку каждого из базовых классификаторов — соответственно, влияние случайностей на усредненную гипотезу существенно уменьшается.
* Он снижает дисперсию. Совокупное мнение целого множества моделей лучше, чем мнение отдельно взятой модели. В экономике это называется диверсификацией — расширение ассортимента выпускаемой продукции повышает эффективность производства и предотвращает банкротство. Ансамбль моделей имеет больший шанс найти глобальный оптимум, поскольку поиск идет из разных точек исходного множества гипотез.
* Он предотвращает выход за пределы множества. Вероятен следующий случай: агрегированная гипотеза находится за пределами множества базовых гипотез. При построении комбинированной гипотезы любым путем (логистическая регрессия, усредненное значение, голосование), множество гипотез расширяется, следовательно, полученный результат не выходит за его рамки.

## 2.1.2 Неконтролируемое обучение

### 2.1.2.1 Алгоритмы кластеризации

Задача кластеризации состоит в группировании множества объектов таким образом, чтобы поместить максимально похожие между собой элементы в одну группу (рис.5).

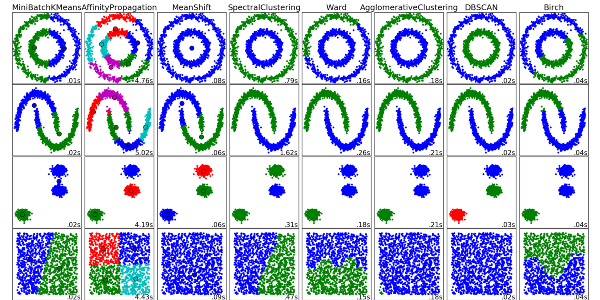


Рисунок 5

Алгоритмов кластеризации существует довольно много, и все они отличаются друг от друга. Самые популярные из них:

* алгоритмы на базе центра тяжести треугольника;
* алгоритмы на основе подключения;
* алгоритмы плотности на основе пространственной кластеризации;
* вероятностный алгоритм;
* алгоритм уменьшения размерности;
* нейронные сети и машинное обучение.

Алгоритмы кластеризации используются в биологии, социологии и информационных технологиях. Например, в биоинформатике с помощью кластеризации анализируются сложные сети взаимодействующих генов, состоящие порой из сотен или даже тысяч элементов. А при анализе результатов социологических исследований рекомендуется осуществлять анализ методом Уорда, при котором внутри кластеров оптимизируется минимальная дисперсия, в итоге создаются группы приблизительно равных размеров.

**2.1.2.2 Метод главных компонент**

Метод главных компонент (PCA) — это статистическая процедура, которая использует ортогональное преобразование с целью конвертации набора наблюдений за возможно коррелированными переменными в набор значений линейно некоррелированных переменных, называемых главными компонентами(рис.6).

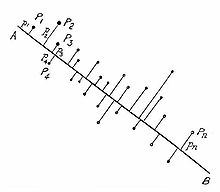


Рисунок 6

Отдельные области применения PCA включают в себя сжатие и упрощение данных для облегчения обучения, а также визуализацию. Решение об использовании метода главных компонент зависит от уровня познания предметной области. PCA не подходит для применения в случаях с плохо упорядоченными данными (все компоненты метода имеют довольно высокую дисперсию).

**2.1.2.3 Сингулярное разложение**

В линейной алгебре под сингулярным разложением (SVD) понимают разложение прямоугольной вещественной или комплексной матрицы. Для матрицы M размерностью [m\*n] существует такое разложение, что M = UΣV, где U и V — унитарные матрицы, а Σ - диагональная матрица.

Метод главных компонент является простым применением сингулярного разложения. Первые алгоритмы компьютерного виденья использовали PCA и SVD, чтобы представить лица в виде суммы базисных компонент, выполнить уменьшение размерности, а затем сопоставить их с изображениями из обучающей выборки. И хотя современные методы характеризуются более сложной реализацией, многие из них по-прежнему работают на базе подобных алгоритмов.

**2.1.2.4 Анализ независимых компонент**

Анализ независимых компонент (ICA) представляет собой статистический метод выявления скрытых факторов, которые лежат в основе множества случайных величин, сигналов и прочих измерений. ICA определяет порождающую модель для исследуемых многофакторных данных, которые обычно подаются в виде большой базы данных образцов. В модели переменные подаются как линейная смесь некоторых скрытых переменных, а любая информация о законах смешивания отсутствует. Предполагается, что скрытые переменные независимы друг от друга и представляются как негауссовские сигналы, поэтому они называются независимыми компонентами исследуемых данных. Анализ независимых компонент непосредственно связан с методом главных компонент, но это гораздо более мощная техника, способная найти скрытые факторы источников, когда классические методы в лице PCA дают сбой. Алгоритм ICA применяется в телекоммуникациях, астрономии, медицине, распознавании речи и изображений, диагностировании и тестировании сложных электронных систем и, наконец, поиске скрытых факторов и источников движения финансовых показателей.

**2.2 Возможности машинного обучения**

#### 2.2.1 Кредитный скоринг

#### Все чаще компании, действующие в сфере кредитования, используют машинное обучения для прогнозирования кредитоспособности клиентов, а также для построения моделей кредитных рисков. Среди таких компаний — Kabbage, Inc., финансирующая малый бизнес посредством платформы кредитования, сервис удаленного микрокредитования LendUp и признанный лидер отрасли финансовых технологий Lending Club. В частности, команда Kabbage специализируется на разработке алгоритмов машинного обучения нового поколения и аналитики для построения моделей кредитного риска и анализа существующего портфеля. В числе множества алгоритмов машинного обучения для определения рейтинга кредитоспособности заемщика используются следующие: многослойный перцептрон, логистическая регрессия, метод опорных векторов, а также алгоритм усиления классификаторов AdaBoost (или Adaptive Boosting) и квантизация векторов при обучении.

**2.2.2 Принятие решений**

Финансовые вычисления и принятие решений могут осуществляться посредством алгоритмов машинного обучения, которые позволяют компьютерам эффективнее и быстрее обрабатывать данные и принимать решения касательно кредитования, страхования, защиты от мошенничества и т. д. Модели машинного обучения широко используются такими компаниями как Affirm, BillGuard и ZestFinance. Последней удалось найти новый подход к традиционным задачам благодаря машинному обучению и анализу больших массивов данных. Компания анализирует тысячи потенциальных кредитных переменных – от финансовой информации до использования технологий, – чтобы лучше оценить такие факторы как возможности потенциального мошенничества, риск невыполнения обязательств и вероятность долгосрочных отношений с клиентами. Как результат, предприятие может принимать более «правильные» решения о предоставлении кредитов, что приводит к повышению доступности кредитов для заемщиков и более высокому проценту их погашения.

**2.2.3 Извлечение информации**

Время поговорить о разновидности информационного поиска, целью которого является автоматическое получение структурированных данных при обработке неструктурированной или слабоструктурированной информации. Как правило, это касается работы с веб-контентом, то есть статьями, публикациями в социальных сетях и различными документами. Например, специализированная система поиска AlphaSense для финансовых компаний использует алгоритмы обработки естественного языка и сложные алгоритмы машинного обучения.

Благодаря мощным алгоритмам собственной разработки фирма Dataminr способна совершить моментальный анализ потоков публикаций Twitter и других данных из социальных сетей и веб-источников и преобразовать их в полезную информацию, которая может быть применена на практике. Компания нацелена на клиентов из сферы финансов и новостей, а также государственного сектора и корпоративной безопасности. Обрабатывая по 500 млн.твитов ежедневно, алгоритмы Dataminr способны отыскать релевантную информацию об интересующих клиентов новых восходящих трендах и последних горячих новостях на минуты и даже часы ранее того момента, когда они, собственно, станут таковыми.

**2.2.4 Защита от мошенничества и управление идентификационной информацией пользователей**

По результатам проведенных IBM исследований, ежегодно мошенничество наносит финансовой индустрии ущерб, равный примерно 80 млрд долларов. Машинное обучение дает более эффективные методы выявления мошенничества. Благодаря созданным решениям можно проводить анализ истории транзакций для построения модели, которая могла бы распознать мошеннические действия. Кроме того, технологии машинного обучения также применяются финтех-компаниями для разработки систем биометрической аутентификации пользователей. Стартап EyeVerify разработал технологию с применением алгоритмов машинного обучения, позволяющую использовать модное «селфи» для обеспечения безопасности своих финансовых операций. Их флагманский продукт Eyeprint ID – программное обеспечение, которое идентифицирует пользователя по рисунку вен на белках глаз и других микроскопических особенностям глаза.

Компания Feedzai, специализирующаяся на обработке и анализе данных, использует машинное обучение и большие массивы данных для повышения уровня безопасности информации предприятий. Разработанные ими модели машинного обучения распознают мошенничество на 30% быстрее, нежели при использовании более традиционных методов обнаружения фрода.

#### 2.2.5 Алгоритмические стратегии для торговли

Машинное обучение применяется для создания высокоэффективных алгоритмических стратегий для торговли. Основной формой алгоритмической торговли является высокочастотный трейдинг, в котором для быстрой торговли ценными бумагами задействованы специальные алгоритмы и торговые роботы. Машинное обучение предоставляет мощные инструменты для изучения закономерностей рынка. Благодаря предиктивному моделированию, программированию и алгоритмам машинного обучения компания KFL Capital Management Ltd., управляющая фондом инвестиций, стала экспертом в области прогнозирования изменений поведения рынка на основании финансовых данных. Торговая фирма Binatix внедряет сверхсовременные алгоритмы машинного обучения, которые помогают обнаружить закономерности, дающие преимущество в инвестировании.

Наряду с развитием других технологий, машинное обучение вносит значительные изменения в финтех-индустрию, предлагая эффективные решения для анализа данных и принятия решений. Алгоритмы машинного обучения используются для многих задач в различных сферах финтеха – от кредитования до повышения безопасности финансовых операций, при этом они могут быть направленны как на индивидуальных клиентов, так и на корпорации.

**ЗАКЛЮЧЕНИЕ**

В ходе написания курсового проекта мною были подробно рассмотрены два основных метода машинного обучения: контролируемое обучение, неконтролируемое обучение, а также возможности машинного обучения такие как:

* Кредитный скоринг
* Принятие решений
* Извлечение информации
* Защита от мошенничества
* Алгоритмические стратегии для торговли

При написании курсовой работы по теме исследования мною была изучена специальная литература, включающая научные статьи по информационным технологиям, учебники по машинному обучению, рассмотрено практическое применение машинного обучения на производстве и бизнесе.

В результате анализа возможностей машинного обучения, мною было выявлено, что сфера применений машинного обучения постоянно расширяется. Повсеместная информатизация приводит к накоплению огромных объёмов данных в науке, производстве, бизнесе, транспорте, здравоохранении. Возникающие при этом задачи прогнозирования, управления и принятия решений часто сводятся к обучению по прецедентам. Раньше, когда таких данных не было, эти задачи либо вообще не ставились, либо решались совершенно другими методами.

С каждым годом большие данные становятся все более сложными и

человеку становится уже недостаточно аналитических

способностей. Некоторые задачи перейдут к искусственному

интеллекту, который справится с ними лучше и быстрее, чем человек.

Прогнозируется, что в ближайшее десятилетие искусственный

интеллект займет около 7% рабочих мест в России. Часть процессов в

организациях автоматизируют и управлять этими процессами поручат

самообучающимся алгоритмам.

Огромное место займет ИИ в интернете вещей. Интернет вещей

требует обработки большого потока информации в реальном времени.

Устройства, подключенные к сети, генерируют гигантские массивы

данных, которые необходимо будет обрабатывать, анализировать и

хранить. Как пример, зубные щетки с функцией bluetooth, которые будут

отправлять информацию о состоянии зубов стоматологу или дроны,

которые станут незаменимы в сельском хозяйстве. Они будут собирать

информацию о зрелости урожая, о состоянии почвы, о вредителях и

болезнях растений. И во всех технологиях будет использовать

машинное обучение. Поэтому машинное обучение готовит для нас

перспективное будущее и массу интересных инноваций.

**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1.Педро Домингос Верховный алгоритм. Как машинное обучение изменит наш мир.-М.: Манн, Иванов и Фербер,2016.-336 с.

2. Флах Питер Машинное обучение.-М.: ДМК-Пресс, 2015 г.-400 с.

3. Бринк Хенрик, Ричардс Джозеф, Феверолф Марк Машинное обучение.-М.: Питер, 2017 г.-336 с.