ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ

РОССИЙСКОЙ ФЕДЕРАЦИИ»

Департамент анализа данных,

принятия решений и финансовых технологий

курсовая РАБОТА ПО ДИСЦИПЛИНЕ «ПРОЕКТНЫЙ ПРАКТИКУМ»

на тему:

**Система искусственного интеллекта на базе нейронных сетей для работы в составе системы лояльности покупателей**

Выполнил:

Студент группы ПИ3-1

Факультета прикладной математики   
и информационных технологий

Матиив Роман Александрович

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Подпись)

Руководитель:

старший преподаватель

Иванов Иван Иванович

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Подпись)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Итоговая оценка)

Москва 2018

Оглавление

[Глоссарий 3](#_Toc514334550)

[Введение 4](#_Toc514334551)

[1. Основная часть 5](#_Toc514334552)

[1.1 Применение карт лояльности «традиционным» способом 5](#_Toc514334553)

[1.2 Применение карт лояльности без активного участия клиента 10](#_Toc514334554)

[1.2.1 Технологии используемые при создании систем автоматического распознавания лиц (face recognition) 10](#_Toc514334555)

[1.2.2 Процессы которые подверглись изменению 16](#_Toc514334556)

[1.3 Техническая реализация данной системы 20](#_Toc514334557)

[1.4 Экономическая обоснованность автоматизации 21](#_Toc514334558)

[Заключение 24](#_Toc514334559)

[Список использованных источник 26](#_Toc514334560)

# **Глоссарий**

* + NN – нейронная сеть
  + CNN – сверточная нейронная сеть
  + Data-driven – развитие или улучшение на основе данных
  + ML – машинное обучение
  + full-connected NN – полно связная нейронная сеть
  + БД – база данных
  + face recognition – распознавание лиц
  + ЭВМ – электронно-вычислительное устройство
  + DS – Dtata Science
  + Бекэнд библиотеки – библиотека которая используется в качестве основы
  + Датасет – набор данных

# **Введение**

Цель данной работы заключается в том, чтобы создать прототип системы которая сможет автоматизировать процесс идентификации пользователей ритейла посредствам NN (если более формально то благодаря CNN). В рамках ограниченного бюджета.

Данная тема особенно актуальна сейчас, когда наблюдается активный рост популярности решений, использующих NN.

Безусловно львиная часть всех решений присутствующих сейчас являются не более чем маркетинговым ходом различных компаний, одним из примеров может служить компания сбербанк, которая утверждает что уже сейчас использует нейронные сети практически во всех своих бизнес юнитах. Однако совершенно очевидно и ясно что выстроить грамотную Data-driven инфраструктуру за 1-2 года очень сложно, из чего можно сделать вывод что не многие компании на сегодняшний день действительно успешно внедряют и используют решений основанные на NN.

В основном это сатрапы или не большие компании, которые могут довольно быстро внедрять какие-либо новые решения без существенной бумажной волокиты и прочих бюрократических аспектов.

Тем не менее абсолютное большинство компаний которые осуществляют свою деятельность в IT (а сейчас фактически все компании имеют IT департаменты) инвестируют в данную сферу, если не NN то как минимум в технологии использующие ML. Поэтому не за горами время когда технологии, активно использующие NN будут активно использоваться. Российский рынок пока до этого не созрел, но я считаю что это вопрос 5-7 лет.

Если говорить про мировые методики решения данной задачи, то в области распознавания изображений на сегодняшний день существенно превалируют CNN.

# **Основная часть**

## **Применение карт лояльности «традиционным» способом**

В современном ритейле карты лояльности это неотъемлемая часть взаимодействия с клиентом. Благодаря ним покупатели получат те или иные скидки, накапливают баллы учувствуют в акциях и тд. С точки зрения бизнеса это отличный инструмент для того, чтобы удержать клиентов, ведь в современном мире конкуренция (особенно в некоторых сферах) просто огромная и заставить клиента не просто посещать, условно, ближайший магазин от дома, но и сделать так, чтобы при возможности выбора он выбирал раз от раза именно определенные магазин это уже довольно не тривиальная задача и карты лояльности это один из таких сильных и устоявшихся инструментов который используют именно в этих целях.

Для того, чтобы понять что нужно автоматизировать, и нужно ли вообще, в этом способе удержания клиента нужно рассмотреть его более подробно.

На Схеме 1 рассмотрен общий процесс взаимодействия с клиентом. Как можно увидеть в таком простом процессе как использование карты лояльности учувствуют 3 актора:

* Покупатель
* Продавец (в частном случае кассир)
* БД

И 2 процесса:

* Получение карты
* Применение карты

Как можно заметить на Схеме 2 процесс получения карты занимает некоторое время, что не являться плюсом как для клиента, который скорее всего не хочет тратить лишнее время на совершение всех действий которые нужно выполнить для получения карты:

1. Получить физическую карту лояльности
   1. Согласиться на предложение продавца
   2. Самому изъявить желание о получении карты
2. Ввести персональные данные
   1. На месте
   2. В другом удобном для клиента место через портал

так и для ритейла, ведь любой из способов получения карты одним клиентом, задерживает всех остальных, что явно не может вызывать у них удовольствие и желание при многократном повторении таких задержек посещать данный магазин.



Схема 1 Взаимодействие ретейла с клиентом



Схема 2 Не автоматизированный процесс получения карты лояльности

Также после того как карта лояльности получена и активирована, для ее успешного применения (Схема 3) клиент должен иметь ее при себе при посещении магазина в каком-либо виде:

* Физический носитель в форм-факторе банковской карты
* В установленном на телефоне приложении
  + Официальном приложении магазина
  + Стороннем агрегаторе карт лояльности

На практике часто случается так, что уже оформленную (часто и не один раз оформленную) карту клиент попросту забывает, теряет или происходит что-то еще благодаря чему он не может ей воспользоваться.

Данная ситуация усугубляется тем, что при отсутствии карты лояльности продавцы очень часто спрашивают о желании оформить последнюю, а учитывая что она уже есть смысла в этом как такого нет, однако слышать данное предложение клиент будет постоянно, что может не нравиться определенным клиентам.

В итоге магазин имеет недовольных покупателей на этапах:

* Получения карты
* Применения карты

Данную проблему безусловно хотелось бы исправить, ведь это логично и интуитивно понятно, что чем больше клиент доволен обслуживанием в магазине, тем больше вероятность того, что он в этот магазин вернётся. И данную вероятность определенно хотелось бы максимизировать. Для того чтобы это сделать необходимо прежде всего формализовать проблему, для того, чтобы найти оптимальный способ ее решения.

Основанная и по сути единственная проблема для клиента заключается в том, что необходимо его активное участие в процессах получения и применения карты. Соответственно для того, чтобы эту проблему исключить нужно исключить и активное участие пользователя их данных процессов, то есть сделать так чтобы оба процесса выполнялись не требуя при этом непосредственного участия пользователь, скрытно он него.



Схема 3 Применение карты лояльности при непосредственном участии клиента

## **Применение карт лояльности без активного участия клиента**

Одним из наилучших решений данной проблемы является использование технологий распознавания образов, а конкретно в нашем случае распознавания образов лица человека (face recognition). Ведь очевидно что лицо у человека всегда при себе, а для того чтобы человека сфотографировать в современном мире не нужно дополнительных усилий со стороны пользователя, делать это можно совершенно не заметно для последнего.

Естественно для создания системы, которая сможет решать проблему постоянной вовлеченности пользователя в процесс требуются технологии и ресурсы которые необходимо знать заранее и понимать нужна ли тому или иному ритейлу данная система.

### **Технологии используемые при создании систем автоматического распознавания лиц (face recognition)**

Решить данную задачу 10-20 лет назад было либо просто невозможно либо экономически неэффективно, было бы выгоднее посадить 50-100 человек с низкой квалификацией, но с хорошим зрением следить за мониторами, которые бы и передавали видео поток с камер, после чего те самые люди уже и опознавали клиентов.

Однако современные технологии позволяют решить данную задачу без привлечения столь большого числа средств и человеческих ресурсов. А все благодаря тому, что люди, в частности ученые, уже достаточно долгое время понимали что человеческий мозг обладаем огромными вычислительными ресурсами, которые к тому же задействуются не на 100% в отличие от компьютеров, которые научились создавать люди. И при всем этом человеческий мозг даже 5 лётного ребенка без труда может распознавать лица, в отличие от обычных алгоритмов, которые могут без труда распознавать лишь форму самого лица, но вот кому конкретно это лицо принадлежит нет.

Однако ученым давно известно, что человеческий мозг состоит из нейронов, поэтому, для того чтобы достигнуть таких же результатов каких достигает человеческий мозг вполне логично попытаться смоделировать его работу. Что непосредственно и было сделано учеными.

На рисунке 1 представлена модель биологического нейрона. Для понимания поставленной перед нами задачи необходимо понимать что :

* Дендрит – так часть нейрона которая получает информацию, входной сигнал.
* Тело нейрона – та часть, которая обрабатывает входные сигналы
* Аксон – часть которая передает обработанные сигналы дальше, в том числе и другим нейронам.

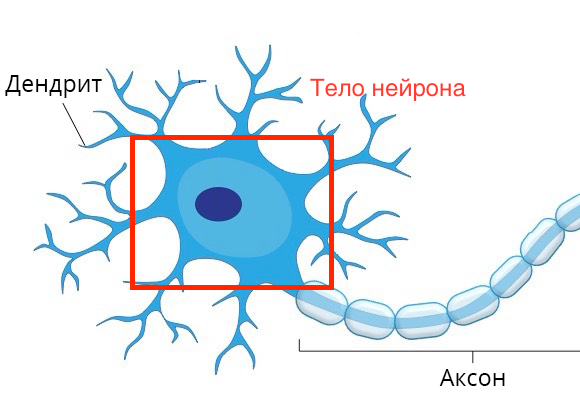


Рисунок 1Модель человеческого нейрона

Но как было упомянуто выше рисунок 1 это биологическая модель нейрона, задача состоит в том, чтобы формализовать все сказанные выше понятия через язык математики, для того, чтобы все это можно было вычислять при помощи современных ЭВМ.

На рисунке 2 представлена более формализованная с точки зрения математики модель биологического нейрона:

* X1 –это 1 сигнал

* – это Дендриты, то что нейрон получает, множество сигналов
* Тело нейрона - та часть, которая обрабатывает дендриты(входные сигналы)
* Аксон - передаёт обработанные сигналы дальше

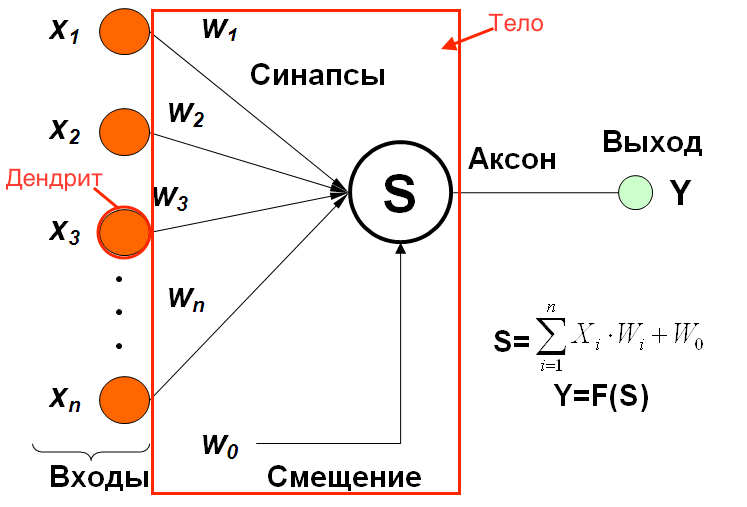


Рисунок 2Математическая модель нейрона

И наконец именно на схеме 4 представлен 1 биологический нейрон, формализованный с точки зрения математики.

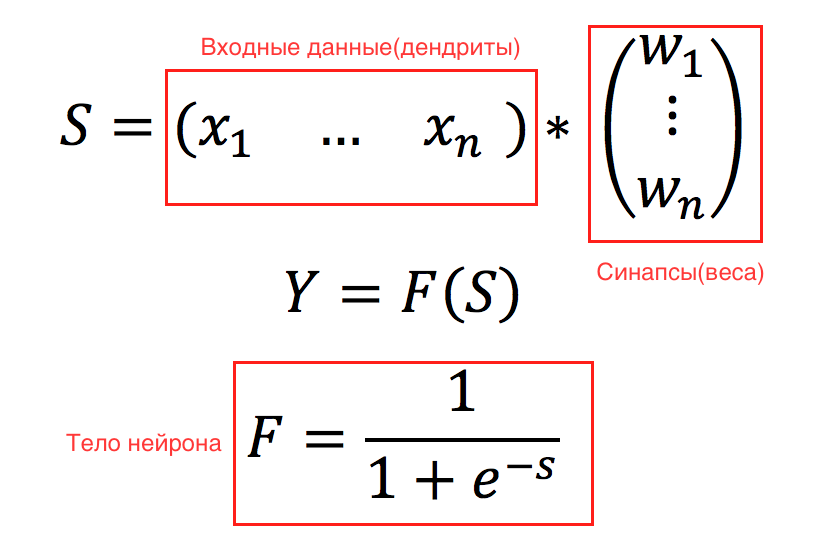


Схема 4

На самом деле многие детали при объяснении данного материала сейчас и далее опущены или изложены очень не формально, например то как правильно подобрать (обучить) синапсы (веса) не сказано ни слова, а также многое другое, но для понимания данной работы информации которая здесь представлена непременно хватит.

Однако, человеческий мозг состоит не из одного нейрона а из совокупности взаимодействующих между собой нейронов, поэтому давайте введем понятие нейронной сети, оно легко обобщается из 1 нейронна.

На схеме 5 пошагово показано как обобщить 1 нейрон до 2 нейронов, после чело до k нейронов (где k-любое натуральное число).

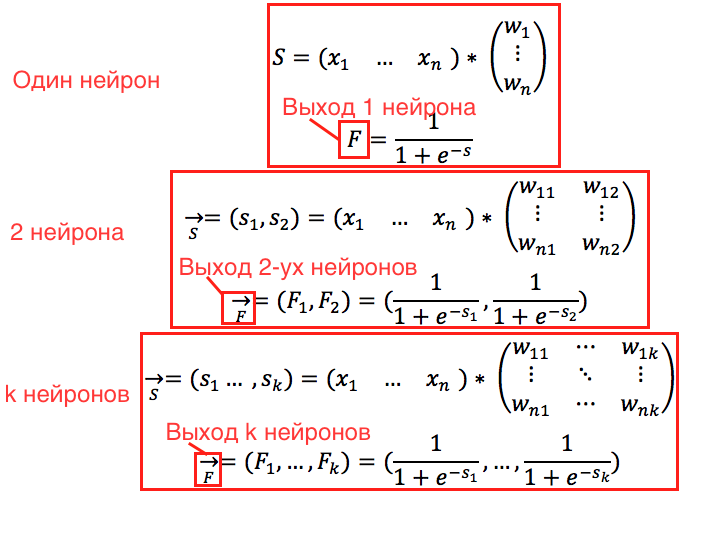


Схема 5

Для полного понимания картины осталось объяснить, что же такое (x1,…,xn). Это собственно и есть наша картинка, а получается она следующим образом:

1. Представим рисунок 3(милого котенка) в виде матрицы 3х3

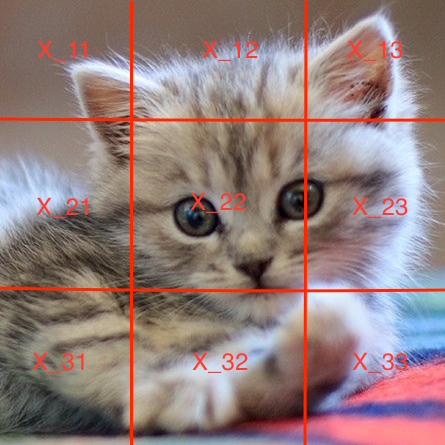


Рисунок 3

1. После чело превратим нашу матрицу 3х3 в вектор 1х9
2. Дальше для дальнейшего удобства перенумеруем пиксели по порядку

Именно так и выглядит преобразование 1 картинки в 1 входной вектор

На схеме 6 представлена простейшая полно связная нейронная сеть, без скрытых слоев. Для лучшего понимания пройдемся по преобразованию шаг за шагом (с лево направо):

1. Картинка преобразованная в вектор матрично умножается на k нейронов, тем самым формируя скрытое представление нашего изображения
2. После этого результат матричного умножения, вектор размерности (1х k, пропускается через нелинейное преобразование (сигмойду) F=. Для каждого полученного значения
3. И наконец последний шаг это предсказание вероятности принадлежности к m классам.

В итоге если m =3 (3 класса) то мы получим 3 числа в диапазоне , которые будут показывать к какому из 3 классов наша картинка вероятнее всего принадледит.

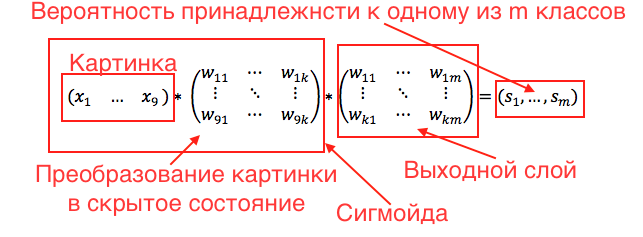


Схема 6

Тем не менее обязательно стоит обратить внимание на то, что материал, приведенный выше предназначен для общего понимания того, как нейронная сеть обрабатывает данные (в нашем случае картинку), представленная выше нейронная сеть имеет архитектуру полно связной (fully-connected).

Однако из практических, и теоретических соображений изображения лучше обрабатывать другой архитектурой, сверочной нейронной сетью (CNN). Объяснение принципов работы данной архитектуры выходит за рамки данной работы. Однако если говорить вкратце, то при использовании CNN изображение не разворачивается из матрицы в вектор, а сохраняет свою структуру, что позволяет существенно улучшить качество моделей основанных на CNN. Именно такую архитектуру я и буду использовать при создании прототипа системы.

### **Процессы которые подверглись изменению**

Общая схема взаимодействия клиента с ритейлером осталась прежней, см схему 1, но вот процесс получения и применения карты изменился значительно.

Процесс получения карты клиентом представлен на схеме 7. Для того чтобы клиент получил карту, теперь не требуется его участие, совсем. Вместо этого его фотографии делаются каждый раз при посещении магазина и при наборе необходимого кол-ва фото (определяется техническим специалистом при проектировании нейросети, в ходе экспериментов) клиенту автоматически выдаётся карта лояльности.



Схема 7 Получение карты лояльности без активного участия

И наконец применения карты лояльности представлено на схеме 8. Данный процесс проходит также абсолютно автономно и не требует никаких действий со стороны клиента. Проходит он в несколько этапов:

1. На кассе делается фото клиента
2. Если он распознан и у него есть карта лояльности, то она используется как в традиционном случае.



Схема 8 Применеие карты лояльности без участия клиента

## **Техническая реализация данной системы**

С технической точки зрения данная задача, как уже было сказано выше, лучше всего решается сверточним нейронными сетями (CNN). Реализовать различного рода сверточную архитектура можно на различных библиотеках например:

* Keras
* TensorFlow
* PyTorch
* Theano
* CNTK

В рамках данной работы нейросеть будет написана на библиотеке Keras. Выбор в сторону данного решения был сделан исходя из того, что библиотека Keras довольно проста для изучения и применения а также в качестве бекэнда использует более низкоуровневые библиотеки, такие как TensorFlow, Theano или CNTK. Однако стоит отметить тот факт, что при необходимости тонкой настройки какого-либо слоя нейросети функционала данной библиотеки может не хватить, но данная проблема может возникнуть тогда, когда качество распознавания должно быть точнее чем у существующих решений. А так как модель предварительно обученная на датасете ImageNet и дообученная на нужных пользователю данных достигает порядка 70% accuracy, то в рамках курсовой работы этого вполне достаточно.

Тем не менее при построении нейросети будут испробаванны не только готовые архитектуры, но и написанные самостоятельно и обученные полностью на нашем датасете. Хотя качество у такой модели вероятнее будет хуже чем у доученной.

Данные для обучения (размеченный датасет с лицами) были взяты из открытого датасета VGG face dataset.

## **Экономическая обоснованность автоматизации**

Очевидно, что любое нововведение требует дополнительных затрат со стороны того, кто эти нововведения внедряет, в нашем случае владелицы ритейлов.

Прежде чем внедрять систему применения карт лояльности без активного участия клиента, нужно прекрасно понимать, сколько это будет приблизительно стоить и в какие сроки может быть сделано, для того, чтобы уже на основе имеющихся данных принимать решения о том насколько такая система необходима, каковы риски при ее внедрении и насколько они критичны.

Для того чтобы рассчитать чистую стоимость выполнения проекта нужно:

1. Понимать какое кол-во времени в человеко-часах займет реализация самого проекта
2. Сколько стоит 1 час работы специалистов, которые компетентны в определенных, необходимых для этого проекта, областях.

В таблице 1 представлены приблизительные расчеты того, сколько может потребоваться исполнителю на реализацию данной системы.

|  |  |  |
| --- | --- | --- |
| Стадия разработки системы | Длительность стадии в человеко-часах | |
| Минимальная | Максимальная |
| Сбор данных | 8 | 32 |
| Выбор нейросетевой архитектуры, ее обучение и проверка | 8 | 32 |
| Повторный добор данных/очистка данных | 0 | 20 |
| Повторное обучение модели | 8 | 16 |
| Внедрение решения на стороне заказчика | 20 | 60 |
| Исправление неисправностей в системе, на стороне заказчика | 0 | 40 |
| Итого | 44 | 200 |

Таблица 1 Расчет времени реализации проекта

В таблице 2 представлена стоимость 1 человеко-часа специалистов из разных областей DS и разных квалификаций.

|  |  |
| --- | --- |
| Специализация | Стоимость в рублях 1 человеко-часа |
| Junior DS | 375 |
| Middle DS | 687 |
| Senior DS | 1562 |
| Junior ML-engineer | 1000 |
| Middle ML-engineer | 1437 |
| Senior ML-engineer | 2187 |
| Junior DLspecialist | 375 |
| Middle DLspecialist | 1250 |
| Senior DLspecialist | 2500 |

Таблица 2 Стоимость человеко-часа разных специалистов

Имея всю необходимую информацию рассчитать чистую стоимость проекта не составит особого труда. Также непременно нужно учитывать и то, что для выполнения разных задач нужны специалисты разного уровня, соответственно поручать простую задачу, которую может сделать junior DS, специалист начального уровня и широкого профиля, более квалифицированному (и более дорогостоящему) не имеет особого смысла.

На таблице 3 представлена конечная стоимость проекта.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Минимальная стоимость | | | Максимальная стоимость | | |
| Челов.-часы | Стоимость 1 часа | Итого | Челов.-часы | Стоимость 1 часа | Итого |
| 8 | 375 | 3000 | 32 | 375 | 12000 |
| 8 | 1250 | 10000 | 32 | 1250 | 40000 |
| 0 | 375 | 0 | 20 | 375 | 7500 |
| 8 | 1250 | 10000 | 16 | 1250 | 20000 |
| 20 | 1000 | 20000 | 60 | 1000 | 60000 |
| 0 | 1000 | 0 | 40 | 1000 | 40000 |
| Итого |  | 43000 |  |  | 179500 |

Получается, что минимальная сумма, которая необходима для выполнения нашей системы составляет 43 000 руб. А максимальная 179 500 руб. Однако ориентироваться на минимальную или максимальную стоимость проекта по отдельности не стоит. На практике стоимость будет варьироваться от одной величины к другой, но к сожалению, как показывает практика, в большинстве своем фактическая стоимость будет смещена в сторону максимальной цены проекта.

# **Заключение**

В процессе выполнения данной курсовой работы, была создана модель, которая с точностью (метрика accuracy) 54,4% распознает человеческое лицо (на отложенной выборке). При том, что модель обучалась распознавать 2607 классов (пользователей), что означает, что она распознает куда лучше случайной (точность у случайной около ). Однако качество несомненно можно улучшить, на мой взгляд до 70%-75% без значительных вложений в вычислительные ресурсы, что и будет мной в дальнейшем сделано.

Непосредственно модель была написана на языке программирования Python с использование высокоуровневой библиотеки Keras, в качестве бекэнда использующую более низкоуровневую библиотеку Tensorflow. Архитектура нейронной сети представлена рисунке 4. Она представляет из себя несколько CNN слоев с батч нормализацией и дропаутом, после чего следуют полно связный слой и последний полно связный слой, который уже непосредственно (используя softmax) предсказывает вероятность того, что фото принадлежит тому или иному человеку.

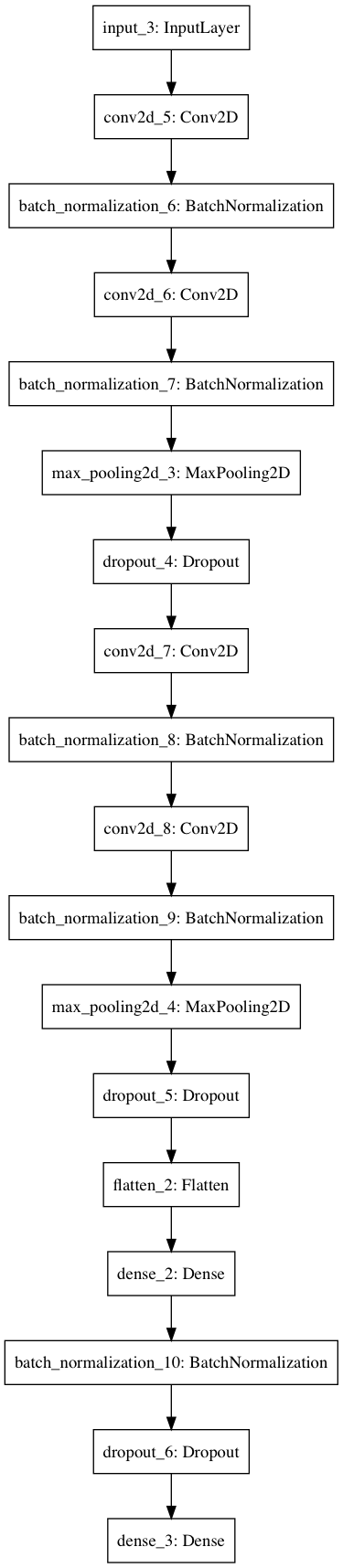


Рисунок 4 Архитектура CNN

# **Список использованных источник**

1. Николенко С. И., Глубокое обучение. СПб.: ПИТЕР, 2018.
2. Ian Goodfellow, Deep Learning: The MIT Press, 2016.
3. Документация библиотеки Keras [Электронный ресурс] – , 2018 – Режим доступа: <https://keras.io> свободный. – Загл. с экрана. – Яз. анг.
4. Stackoverflow информационный портал [Электронный ресурс] – Режим доступа: https://ru.stackoverflow.com свободный. – Загл. с экрана. – Яз. анг.
5. Arxix информационный портал [Электронный ресурс] – Режим доступа: <https://arxiv.org/pdf/1512.03385.pdf> свободный. – Загл. с экрана. – Яз. анг.
6. Arxix информационный портал [Электронный ресурс] – Режим доступа: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> свободный. – Загл. с экрана. – Яз. анг.
7. Stanford vision lab информационный портал [Электронный ресурс] – Режим доступа: <http://cs231n.stanford.edu/2017/syllabus.html> свободный. – Загл. с экрана. – Яз. анг.
8. Stanford vision lab информационный портал [Электронный ресурс] – Режим доступа: <http://cs231n.github.io/convolutional-networks/> свободный. – Загл. с экрана. – Яз. анг.
9. Stanford vision lab информационный портал [Электронный ресурс] – Режим доступа: <http://cs231n.github.io/neural-networks-1/> свободный. – Загл. с экрана. – Яз. анг.
10. Stanford vision lab информационный портал [Электронный ресурс] – Режим доступа: <http://cs231n.github.io/neural-networks-2/> свободный. – Загл. с экрана. – Яз. анг.
11. Stanford vision lab информационный портал [Электронный ресурс] – Режим доступа: [http://cs231n.github.io/neural-networks-3/](http://cs231n.github.io/neural-networks-2/) свободный. – Загл. с экрана. – Яз. анг.
12. C. M. Bishop. Neural networks for pattern recognition. Oxford university press, 1995.
13. I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks. arXiv:1302.4389, 2013.
14. B. D. Ripley. Pattern recognition and neural networks. Cambridge university press, 1996.
15. Andrew G. Howard. Some improvements on deep convolutional neural network based image classification. CoRR, abs/1312.5402, 2013.
16. Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25, pages 1106-1114, 2012.
17. Thomas Serre, Lior Wolf, Stanley M. Bileschi, Maximilian Riesenhuber, and Tomaso Poggio. Robust object recognition with cortex-like mechanisms. IEEE Trans. Pattern Anal. Mach. Intell., 29(3):411-426, 2007.
18. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.
19. R. Szeliski. Locally adapted hierarchical basis preconditioning. In SIGGRAPH, 2006.
20. M.D.ZeilerandR.Fergus. Visualizing and understanding convolutional neural networks. In ECCV, 2014.