**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**

**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук

Образовательная программа «Прикладная математика и информатика»

**Отчет о программном проекте**

на тему "Обучающие материалы отечественной нейросетевой платформы PuzzleLib"

(промежуточный, этап 1)

**Выполнил**:

студент группы БПМИ 198 В. В. Вьялков



Подпись И.О. Фамилия

08.02.2021

Дата

**Принял**:

руководитель проекта Дмитрий Валерьевич Пантюхин \_

Имя, Отчество, Фамилия

Старший преподаватель \_

Должность, ученое звание

Департамент программной инженерии \_

Место работы (Компания или подразделение НИУ ВШЭ)

|  |  |  |
| --- | --- | --- |
| Дата проверки 2021 |  |  |
|  | Оценка | Подпись |
|  | (по 10-тибалльной шкале) |  |

**Москва 2021**

# Содержание

[Содержание 2](#_Toc73454282)

[Основные термины и определения 3](#_Toc73454283)

[Введение 5](#_Toc73454284)

[Сравнительные анализ источников и аналогов 7](#_Toc73454285)

[Описание требований к программному проекту 8](#_Toc73454286)

[Список источников 9](#_Toc73454287)

# Основные термины и определения

**Программирование** — раздел информатики, область знаний об алгоритмах и программах и их свойствах, а такжеисполнителях алгоритмов и программ

**Машинное обучение (Machine Learning)** - класс методов, характерной чертой которых является не прямое решение задачи (на основе правил / формул), а обучение в процессе применения решений множества сходных задач.

**Нейронная сеть (Neural Network)** — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.

**Библиотека** – набор готовых функций, которые позволяют нам собрать и обучить нейронную сеть без необходимости глубоко понимать математическую базу, лежащую в основе этих процессов

**Архитектура сети** - совокупность значимых параметров сети, определяющих её назначение и возможности, среди которых: структура и конфигурация слоев нейронной сети а также характер взаимодействия между слоями.

**Google Colaboratory** - облачная платформа с предустановленным программным обеспечением и библиотеками для проведения вычислений, которую удобно использовать в том числе для обучения глубоких нейронных сетей.

**Jupyter Notebook** — это командная оболочка для интерактивных вычислений, позволяющая создавать файлы с разметкой и кодом. Аналог Google Colaboratory

**Пайплайн обучения –** последовательные стадии подготовки и преобразования данных перед загрузкой в модель

# Введение

С появлением программирования человечество научилось решать множество ранее нерешаемых задач. Знания об алгоритмах, языки программирования и сами комьютеры (с постоянным увеличением вычислительных мощностей), позволили оптимизировать множество процессов.

Прорывным направлением программирования в наши дни является использование машинного обучения для решения прикладных задач. Первые нейронные сети, появились еще в XX веке, а в XXI темпы развития этой отрасли науки значительно ускорились. Сейчас нейронные сети способны решать задачи прогнозирования, классификации, оптимизации, генерации и многие другие.

Программирование нейронных сетей зачастую предполагает использование готовых библиотек с заранее определенными модулями. В наше время существует большое количество инструментов для машинного обучения (Tensor Flow, Keras, Torch и др). Различия между ними прежде всего определяется количеством доступных модулей, быстродействием полученных программ и удобством использования библиотеки программистами.

Библиотека PuzzleLib – набор программных модулей для использования в машинном обучении, полностью разработанный в России. В некоторых задачах этот инструмент превосходит по быстродйествию зарубежные аналоги, но проигрывает в количестве функционала и понятности документации. Ключевыми задачами компании, отвечающей за разработку и поддержку этой библиотеки, являются дополнение библиотеки новыми модулями и написание понятной документации. Мотивация данного проекта происходит из задач компании.

**Цель проекта:** разработка понятной документации для библиотеки Puzzlelib и дополнение функционала новыми модулями

**Задачи проекта:**

* Изучить методы машинного обучения и их применение для решения прикладных задач
* Проанализировать документацию PuzzleLib, выявить слабые стороны, разработать методы для повешния удобства использования библиотеки через переработку и дополнение документации
* Проанализировать фукнционал библиотеки и предложить модули для его расширения
* Разработать новые модули для библиотеки PuzzleLib
* Провести переработку библиотеки, повысить читаемость, исравить ошибки, сделать ее более понятной

# Сравнительные анализ источников и аналогов

В качестве источников для реализации задач проекта были выбраны платформа Coursera, несколько специализированных сайтов по машинному обучению и сайты с документацей зарубежных аналогов. (см. Список источников).

Стоит отметить, что на платформе Coursera, курсы «Введение в машинное обучение» и «Машинное обучение и анализ данных» понятным языком объясняют устройство нейронных сетей. Этот ресурс поможет в освоении машинного обучения прежде всего тем, кто до этого не сталкивался с этой отраслью.

Из аналогов библиотеки PuzzleLib стоит выделить только зарубежные, разработок подобной сложности в России найдено не было. Рассмотрим ключевые аналоги: TensorFlow, Keras, Pytorch.

* Наличие и качество документации
* Быстродействие
* Качество поддержки

|  |  |  |  |
| --- | --- | --- | --- |
| **Критерий сравнения** | **TensorFlow** | **Keras** | **Pytorch** |
| Наличие и качество документации | Компнаия Google снабдила библиотеку понятной документацией, но на уровне тех, кто уже хорошо понимает в разработке | Средней понятности документация, похоже на TensorFlow, есть гайды с разбором конкретных задач | Самая понятная и подробная документация из представленных |
| Быстродействие (По приоритету) | 2 | 3 | 1 |
| Качество подержки | Хорошее, большое количество материвалов информации на форумах | Хорошее, большое количество материвалов информации на форумах, большое комьюнити | Большое комьюнити, на StackOverflow разрабрано много проблем |

# Описание требований к программному проекту

В рамках группового проекта задачи были поделены на несколько частей. Перечень задач, глобально рассмотренных в требованиях к программному проекту, непосредственно закрепленных за мной, ограничивается следующими:

- Написание документации для некоторых частей раздела Modules и Optimizers

- Реализация архитектуры и пайплайна обучения сети DenseNet

- Реализация архитектуры и пайплайна обучения одного из типов сетей GAN

**Функциональные требования**

* Взаимодействие пользователя с программным кодом (Интерактивное)
* Реализация архитектур и пайплайна обучения
* Реализация доступных и актуальных для прикладных задач примеров для библиотеки
* Программный код должен быть снабжен подробными комментариями на русском языке
* Программный код должен быть совместим с библиотекой PuzzleLib

**Нефункциональные требования**

* Взаимодействие пользователя с программным кодом реализовано через Google Colaboratory или Jupyter Notebook
* Документация должна быть написана без ошибок, понятным и доступным для неподготовленного пользователя языком
* Во время исполнения кода из примеров не должно возникать багов

# Список источников

1. Pytorh Documentation [Электронный ресурс]. <https://pytorch.org/docs/stable/index.html> Режим доступа: Свободный (дата обращения: 22.12.2020)
2. TensorFlow Documentation [Электронный ресурс]. <https://www.tensorflow.org/api_docs> Режим доступа: Свободный (дата обращения: 22.12.2020)
3. PuzzleLib Documentation [Электронный ресурс].   
   <https://puzzlelib.org/ru/> Режим доступа: Свободный (дата обращения: 22.12.2020)
4. Keras Documentation [Электронный ресурс].   
   <https://keras.io/> Режим доступа: Свободный (дата обращения: 22.12.2020)
5. Coursera, Введение в машинное обучение [Электронный ресурс].   
   <https://ru.coursera.org/learn/vvedenie-mashinnoe-obuchenie> Режим доступа: Ограниченный (дата обращения: 22.12.2020)
6. Coursera, Машинное обучение и анализ данных [Электронный ресурс].   
   <https://ru.coursera.org/specializations/machine-learning-data-analysis> Режим доступа: Ограниченный (дата обращения: 22.12.2020)
7. Habr, Сравнение фреймворков для машинного обучения [Электронный ресурс]  
   <https://habr.com/ru/company/otus/blog/443874/> Режим доступа: Свободный (дата обращения: 22.12.2020)
8. Термины и опредления в машинном обучении [Электронный ресурс]  
   <https://neural-university.ru/neural-networks-basics#rec116464826> Режим доступа: Cвободный (дата обращения: 22.12.2020)