Оглавление

[Введение 2](#_Toc471722376)

[Средства реализации проекта 3](#_Toc471722377)

[Набор данных 4](#_Toc471722378)

[Работа с данными 5](#_Toc471722379)

[Методы машинного обучения 6](#_Toc471722380)

[Вывод 7](#_Toc471722381)

# Введение

Проект представляет собой сервис по подбору фильмов, реализация которого представлена в форме телеграмм-бота, который отвечает на определённые запросы пользователей.

Для начала проводится анализ присланного боту фото, на котором отыскивается лицо и предугадывается набор эмоций, которые возможно испытывает человек на фото. Затем выделяется основная эмоция, которая преобладает в данный момент и по ней производиться дальнейший подбор фильма. Также в качестве обучения бота пользователю выдается второй фильм на оценку, тем самым на сервере выстраивается весовой граф, где мерой веса фильма изначально является рейтинг, изменяемый впоследствии пользователем.

# Средства реализации проекта

Проект реализован с помощью комплекса технологий на языках C# и PHP. В частности, на C# написан бот и определитель эмоции, которая затем передается на PHP сервер, содержащий dataset из фильмов, который подбирает фильм для просмотра.

Этот стек технологий выбран т.к. и для технологий .NET и для языка PHP существует набор библиотек машинного обучения, а также в виду того что для постоянного пополнения dataset’а можно использовать PHP парсер. Этот же парсер использовался для начального наполнения базы.

Для хранения базы данных использовалась СУБД MYSQL.

# Набор данных

В качестве набора данных использовались данные о 10000 фильмах с сайта imhonet.ru, в частности в базе хранятся следующие поля:

* ID фильма – идентификатор фильма на сайте.
* Title – название фильма, используется при отправке фильма пользователю.
* Genre – жанр фильма, на котором частично основан подбор, т.к. вполне возможно что пользователю с определенной эмоцией интересен фильм определенного жанра. В случае, когда эмоция нейтральна выдается случайный фильм. В дальнейшем вес фильма для каждой эмоции меняется, тем самым дополнительно судить какой жанр более приемлем для определенной эмоции.
* Description – описание фильма, используется при отправке и выводе фильма на стороне пользователя.
* Rating – рейтинга фильма по версии IMDB, т.к. на сайте содержится информация так же о собственном рейтинге фильма и рейтинге кинопоиска то так же можно выстраивать весовые графы и по ним. Является мерой веса фильма, в начале для всех эмоций выставляется общий рейтинг, затем он меняется в зависимости от голоса пользователя.

Единственная информация, хранимая о пользователе это его ID, которое используется для идентификации диалога ботом. Хранится и отдается телеграммом.

# Работа с данными

Работа с данными сводится к тому, чтобы предугадать какой фильм понравится пользователю в данный момент, и соответственно выдать пользователю этот фильм.

Основной алгоритм начинается с того, чтобы при получении фотографии, бот произвёл анализ этого фото на присутствие лиц. И если фото удовлетворяет условию, т.е. на фотографии присутствует как минимум одно лицо, то начать анализ каждого лица по отдельности, и составить небольшие карточки о каждом найденном человеке на фото, которые будут включать его пол, примерный возраст, и основную эмоцию, которую в данный момент человек испытывает больше всего.

Теперь имея примерную картину людей на фото, мы можем произвести подбор. Т.к. в начале сеть является не обученной и обучение производится самими пользователями то фильм выдается в определенном жанре с высоким рейтингом, в дальнейшем весы корректируются и подбор идет по более точным данным.

# Методы машинного обучения

Примеров применения машинного обучения здесь несколько. Во-первых, оно применяется при распознавании лиц. Во-вторых, при использовании методов определении эмоций. Так же строится весовой граф для подбора фильмов.

Как говорилось ранее, человеку отправляется случайный известный фильм, который человек должен оценить. После некоторого количества полученных данных, можно определить примерную оценку любого другого фильма. Как известно, фильм очень редко имеет один жанр, и поэтому сходство фильмов можно определить по максимальному количеству совпадения жанров, так же и по нескольким другим критериям. Именно по этой причине можно говорить о машинном обучении. И если говорить о рейтинге как о определённом классе фильма, вполне можно использовать классификатор *k Nearest Neighbor* для определения рейтинга фильма.

# Вывод

Методы машинного обучения применимы везде. И их польза в современном мире неоспорима.

В нашем случае, как мы надеемся удалось создать вполне хорошее приложение по подбору фильмов для пользователя при помощи определения его эмоций. И применение машинного обучения здесь крайне обширно, от распознавания лиц и чтения эмоций, до применения нейросети для качественного отбора фильмов по эмоциям.

Одно из преимуществ в том что подобным образом можно производить подбор по эмоциям не только фильма, но и других предметов/данных, т.к. уже имеется определитель эмоции по фото, и достаточно будет провести анализ и сопоставить эмоция новый предмет/данные.

Это объясняет тот факт, что возможностей применения машинного обучения очень велико в любой технологии, как и велико, если даже не бесконечно, количество технологий к которым применимо машинное обучение