

## НЕЙРОСЕТЕВАЯ СИСТЕМА ОЦЕНКИ ЗАВИСИМОСТИ ЦВЕТОВ ФОТОГРАФИЙ ПРОФИЛЯ В СОЦИАЛЬНОЙ СЕТИ И КОЛИЧЕСТВА ОТМЕТОК «НРАВИТСЯ»

Шарилова А. О.

Национальный исследовательский университет «Высшая школа экономики»,  
г. Пермь, Российская Федерация

Научный руководитель: Шестакова Л. В. – канд. физ.-мат. наук., доцент кафедры информационных технологий в бизнесе

**Аннотация.** В статье рассмотрена степень текущей разработанности темы продвижения профилей в социальной сети с помощью алгоритмов машинного обучения. Также описано получение и подготовка данных для обучения нейронной сети, проведено её обучение, проверено качество обучения сети. После этого выявляется закономерность, имеющее практическое значение, а также представлен прогноз нейронной сети количества отметок «нравится» для нескольких новых фотографий.

**Ключевые слова:** искусственный интеллект, нейронная сеть, социальные сети, продвижение, анализ фотографий.

**Введение.** В современном мире трудно представить существование малого бизнеса без присутствия в сети Интернет, в социальных сетях. В частности, как показало исследование Крыловой Д. С., малый и средний бизнес всё чаще выбирает в качестве канала коммуникации с целевой аудиторией, а также в качестве способа продвижения социальные сети. И в том числе этим обуславливается рост компаний [1].

Однако одного присутствия в социальных сетях недостаточно для роста бизнеса. Представители малого бизнеса, индивидуальные предприниматели вынуждены или нанимать специалиста по продвижению в социальных сетях, или изучать этот вопрос сами.

В последнем случае предпринимателям будет полезно иметь основу для своего продвижения. В частности, иметь результаты анализа собственного профиля для того, чтобы сделать выводы для последующих публикаций. Данное исследование сконцентрировано на анализе визуальной составляющей профилей социальной сети «ВКонтакте» – фотографий. В работе будет оценена зависимость между цветами фотографий и количеством у них отметок «нравится».

Предполагается обучение модели на наборе фотографий, выявление наиболее удачного для фотографий цвета, а также оценка количества отметок «нравится» у фотографий, не вошедших в обучающее и валидирующее множества.

**Основная часть.** Как было сказано выше, данное исследование заключаться в анализе фотографий профилей представителей малого бизнеса «ВКонтакте». Для получения этой информации было использовано *API (application programming interface)* «ВКонтакте». Далее для определения цветов фотографий были использованы возможности *Azure Cognitive Services* Компьютерное зрение. После этого необходимо было сформировать таблицу из полученных результатов, для этого была использована библиотека *pandas* языка *python*. Далее каждый цвет фотографии был вынесен в отдельный столбец и закодирован методом *One-Hot Encoding*. Полученная таблица была выгружена в *csv* файл.

Полученный набор данных содержал около 1100 строк и был разделён на обучающую, валидационную и тестовую выборки.

Для построения модели были получены следующие параметры: X1 – чёрный цвет; X2 – синий цвет; X3 – коричневый цвет; X4 – зелёный цвет; X5 – серый цвет; X6 – оранжевый цвет; X7 – розовый цвет; X8 – фиолетовый цвет; X9 – красный цвет; X10 – сине-зелёный цвет; X11 – белый цвет; X12 – жёлтый цвет. Выходной параметр – количество отметок «нравится».

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [2] по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет двенадцать входных нейронов, один выходной и один скрытый слой с 38 нейронами на нём.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 58 примеров. Средняя относительная ошибка тестирования составила 35%, что не является приемлемым результатом даже для данных из социальной сети. Следовательно, требуется либо доработка модели, либо чистка исходного набора данных поиск выбросов в нём.

При детальном рассмотрении исходного набора данных оказалось, что в нём присутствуют «противоречащие» примеры, связанные с особенностью социальной сети. В частности, очень похожие фотографии из одной серии, имеющие одинаковые цвета отмечены разным количеством «лайков» (рисунок 1).

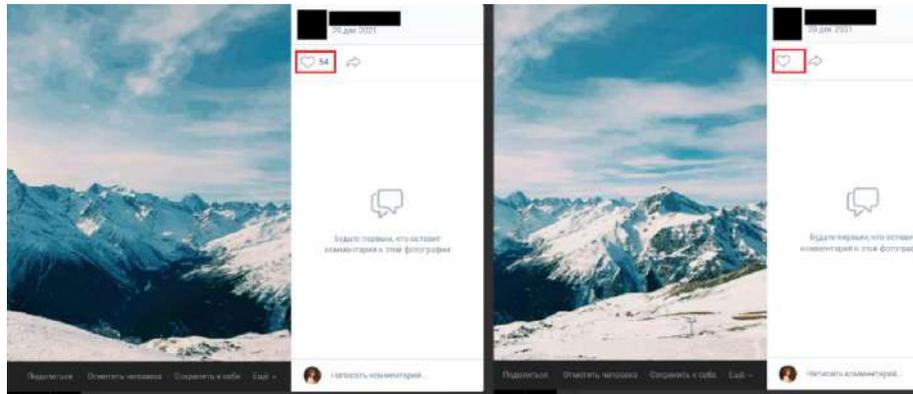


Рисунок 1 – Почти одинаковые фотографии с разным количеством «лайков»

После выявления таких «дубликатов» из всех множеств были удалены примеры с нулевым количеством отметок «нравится». Также было решено немного упростить сеть уменьшив количество нейронов на скрытом слое до 13 и установить линейную активационную функцию выходного слоя.

Была построена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью количеством отметок «нравится» (рисунок 2). Средняя относительная ошибка тестирования составила 27,8%. Расчёт среднеквадратичной ошибки в нейросимуляторе производится с помощью следующей формулы:

$$E = \frac{\sqrt{\frac{\sum_{n=1}^N (d_n - y_n)^2}{N}}}{|\max(d_n) - \min(d_n)|} 100\% , \quad (1)$$

где  $N$  - количество элементов выборки,

$d_n$  - фактическое значение моделируемой переменной,

$y_n$  – значение переменной, вычисленное с помощью нейронной сети.

Данная средняя квадратичная относительная ошибка тестирования нейронной сети является допустимой для данных, собранных вручную из социальной сети.

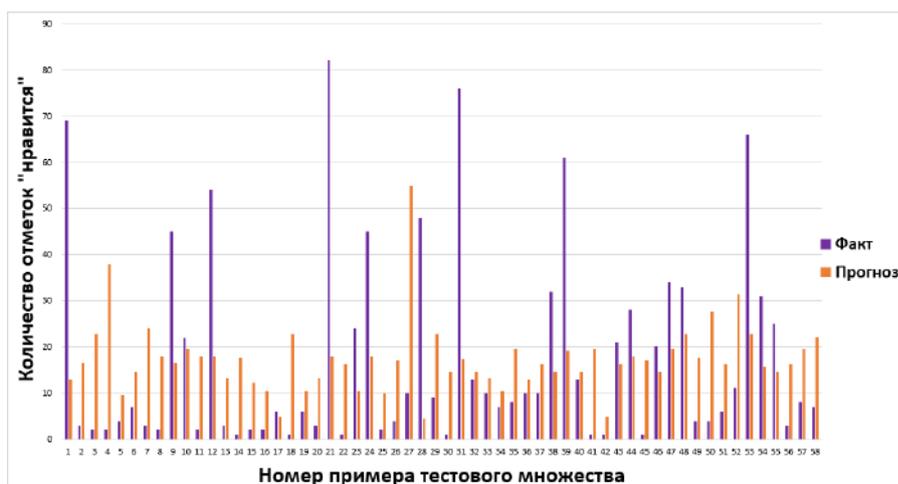


Рисунок 2 – Результат тестирования нейронной сети

Присмотревшись к полученному графику, можно заметить, что присутствуют отклонения в предсказаниях нейронной сети. Это подтверждается и коэффициентом детерминации, который оказался меньше 0,3. В будущем планируется уменьшение таких отклонений с более глубокой работой с исходными данными для обучения нейронной сети. Например, планируется строить отдельные модели для отдельных профилей. Так как у каждого профиля своя аудитория и предпочтения, соответственно есть основания для проектирования и обучения сети на фотографиях конкретного пользователя.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5». Была оценена значимость на обучающем и на тестовом множествах, далее был вычислен средний результат. Оказалось, наиболее значимым цветом на фотографиях является жёлтый. Теперь необходимо понять положительно ли влияет наличие жёлтого цвета на фотографии или отрицательно.

Для этого был применён метод сценарного прогнозирования [4-8]. Необходимо было рассмотреть ранее неопубликованные фотографии, которые не входили в обучающее, тестовое и валидирующее множества (рисунок 3).

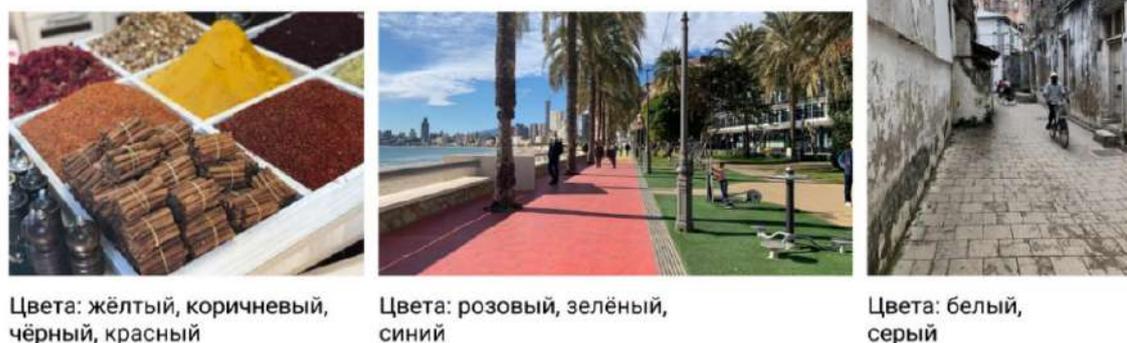


Рисунок 3 – Фотографии и их цвета для прогноза

В Нейросимулятор были внесены следующие записи для прогноза с ожидаемым результатом (таблица 1).

Записи были занесены в Нейросимулятор и были построены прогнозы отметок «нравится». Как оказалось, ожидаемые результаты полностью совпали с прогнозом Нейросимулятора. Жёлтый цвет действительно определяет высокое количество отметок «нравится» у первой фотографии и его присутствие может повысить количество «лайков» второй и третьей фотографии на более чем 50%. Получается, что яркие фотографии с

присутствием жёлтого цвета самые выигрышные для публикации на странице в социальной сети. Это может быть объяснено тем, что психологически жёлтый цвет у людей ассоциируется с солнцем, радостью.

Таблица 1 – Цвета фотографий для прогноза и ожидаемый результат

Фото	Цвета	Ожидаемый результат
1	Жёлтый, коричневый, чёрный, красный	Высокое количество «лайков»
1	Коричневый, чёрный, красный	Количество «лайков» ниже, чем в первом случае
2	Розовый, зелёный, синий	Низкое количество «лайков»
2	Розовый, зелёный, синий, жёлтый	Количество «лайков» выше, чем в первом случае
3	Белый, серый	Низкое количество «лайков»
3	Белый, серый, жёлтый	Количество «лайков» выше, чем в первом случае

**Заключение.** В рамках данной работы была построена система нейросетевого прогнозирования количества отметок «нравится» для цветов фотографий профиля социальной сети. Спроектированная нейросетевая модель учитывает 12 цветов: чёрный, синий, коричневый, зелёный, серый, оранжевый, розовый, фиолетовый, красный, сине-зелёный, белый и жёлтый цвета. Оказалось, что наиболее всего на высокое количество «лайков» влияет присутствие жёлтого цвета на фотографиях.

### Список литературы

1. Крылова Дарья Сергеевна Влияние социальных сетей на рост компаний малого и среднего бизнеса // Стратегии бизнеса. 2017
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
3. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
4. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. 2013. № 6. С. 985.
5. Ясницкий Л.Н., Уткова О.В. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 1. С. 189.
6. Ясницкий Л.Н., Кировова А.В., Ратегова А.В., Черепанов Ф.М. Прогноз результатов чемпионата мира-2015 по легкой атлетике методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 4. С. 624.
7. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

UDC 004.032.26

## NEURAL NETWORK SYSTEM FOR ASSESSING THE DEPENDENCE OF THE COLORS OF A PROFILE PHOTOS IN A SOCIAL NETWORK AND THE NUMBER OF LIKES

Sharipova A. O.

National Research University Higher School of Economics, Perm, Russian Federation

Shestakova L.V. – PhD, assistant professor, associate professor of the department of ITIB

**Annotation.** The article considers the degree of current development of promoting profiles in a social network using machine learning algorithms. It also describes the acquisition and preparation of data for training a neural network and how its training was carried out. Then, the quality of network training was checked. After that, a pattern of practical importance is revealed, and a neural network forecast of the number of “likes” for several new photos is also presented.

**Keywords:** artificial intelligence, neural network, social networks, promotion, photo analysis.