

Анализ психоэмоциональных образов клиентов для оказания консалтинговых услуг банкам

Мазуров Павел Сергеевич
РЭУ им. Г.В. Плеханова
студент

Аннотация

В статье описано понятие «скоринга», даны характеристики данного процесса. Выделяются те характеристики, которые наиболее тесно связаны с ненадежностью или, наоборот, с надежностью клиента. Содержатся основные виды данных для моделей скоринга. Цель: описать и проанализировать процесс скоринга, а также его основные характеристики; рассмотреть исследования с использованием механизмов скоринга. Материалы и методы: результаты опросов пользователей социальных сетей, исследования различных университетов в данной области.

Ключевые слова: скоринг, скоринг-модели, кредитование, дихотомические переменные, социальные сети, лайки, прогнозирование данных, психометрические данные

The analysis of the psychoemotional images of clients for providing the consulting services to the banks

Mazurov Pavel
Plekhanov Russian University of Economic
student

Abstract

The article describes the definition of “scoring”, the characteristics of this process. There are characteristics those are the most closely related to unreliability or, conversely, to the reliability of the client. Main types of data for scoring models are contained. Purpose: to describe and analyze the scoring process, as well as its main characteristics; Consider research using scoring mechanisms. Materials and methods: results of surveys of users of social networks, studies of various universities in this area.

Key words: scoring, score models, lending, dichotomous variables, «social networks, likes, data prediction, psychometric data

Введение

Основным условием эффективной деятельности коммерческих банков на рынке кредитных услуг является применение методов и моделей объективной оценки кредитоспособности заемщиков, позволяющих банкам принимать экономически обоснованные решения и выбирать наиболее

приемлемые альтернативы размещения кредитных ресурсов банка [2]. В последние годы значительно возросло число публикаций по проблемам управления кредитным портфелем, принятию кредитных решений в условиях риска и неопределенности. Тем не менее данный вопрос все еще остается недостаточно освещенным. Теоретическое обоснование и практика разработки моделей оценки кредитоспособности предприятий-заемщиков в России еще недостаточны, а приложимость западного опыта к российской действительности вызывает серьезные затруднения. Таким образом, существует объективная необходимость проведения комплексных исследований по обозначенной проблеме.

Целью данной работы является описание методики построения рейтинговых скоринг-моделей оценки кредитоспособности предприятий-заемщиков.

Актуальность данной темы обусловлена тем, что в настоящее время сформировано множество банков. Они образовали большую банковскую сеть. Цель банков – повышение доходности, а повышение доходности зависит от управления кредитными рисками. Снизить риски без потерь доходности поможет скоринговая система.

Скоринг - метод присвоения численного значения (“счета”) клиенту, с целью предсказать, с какой вероятностью клиент или клиентка банка столкнется с определенным случаем или ситуацией в будущем относительно других. Это основано на представлении о том, что поведение прошлых лет является основой поведения населения в будущем, имея аналогичные характеристики.

При анализе образцов данных истории клиента, а также бизнес-данных, выводятся тенденции для лучшего понимания клиентов (в т. ч. потенциальных клиентов) и прогнозирования последующих событий, например, выплату кредита.

В механизме анализа клиентов основное внимание уделяется кредитным рейтингам. Они представляют собой числовые данные, оценивающие кредитоспособность клиента или вероятность того, что этот клиент погасит кредит. Финансовые институты используют оценочные модели для определения кредитного риска заемщика и оказания помощи во время оценки кредита.

Оценка может быть применена на этапах методологии кредитования микрофинансирования. Это может обеспечить объективность для того, чтобы сделать процесс более эффективным, а также улучшить стандарты и средства контроля. Кредитная оценка модели не предназначена для замены банковских служащих и другого коммерческого персонала. Скорее, она создана для дополнения и облегчения их работы [8].

Исходные данные скоринга

Модели Скоринга анализируют данные клиента, чтобы определить тенденции, которые могут помочь предсказать будущее поведение.

Существуют различные источники данных, которые счет может использовать в качестве входа:

1. Социально-демографические данные, такие как, например, пол клиента, возраст, гражданский статус, деловой опыт, длительность проживания в текущем месте жительства, тип места жительства, и т.д.

2. Поведенческие данные, которые отражают историю отношений клиента с учреждением и могут включать следующие переменные: максимальное количество дней, в течение которых клиент может просрочить выплату кредита; среднее количество дней, в течение которых клиент не выполнил свои обязательства; количество случаев, когда клиент просрочил выплату кредита на 1 день и т. д.

3. Данные кредитного агентства, которые собраны третьим лицом и включают личную кредитную историю с финансовыми учреждениями.

В последние годы микрофинансовые учреждения в сотрудничестве с третьими лицами также исследовали более новые, дополнительные источники данных, включающие в механизмы управления рисками такие Скоринг-модели:

1. Психометрические данные прибывают из специально предназначенных анкетных опросов, чтобы определить знание человека, способности, отношения и черты индивидуальности. В микрофинансах эти принципы могут использоваться для рейтинга кредитоспособности, а также могут оценить предпринимательские способности.

2. Большие данные включают данные из различных источников за пределами учреждения, таких как телефонные компании, утилиты, ритейлеры или социальные сети (Facebook, Твиттер, LinkedIn, Yahoo!). [2]

Одним из инструментов скоринга, который появился относительно недавно, можно назвать приложение myPersonality.[1]

myPersonality - популярное приложение в Facebook, позволяющее пользователям проходить реальные психометрические тесты и записывающее (с их согласия) свои психологические данные и профили Facebook. В настоящее время база данных приложения содержит более 6 000 000 результатов тестов, а также более 4 000 000 индивидуальных профилей Facebook. Респонденты находятся в разных возрастных группах, имеют разное происхождение и культуру. У них большая мотивация для честных ответов, поскольку единственное, что они получают за свое участие - это отчет об их результатах.

Зарегистрированным сотрудникам доступны самые разнообразные данные, в том числе:

- оценки психометрических тестов;
- записи профилей пользователей в Facebook;
- данные уровня предмета.

Это приложение позволяет оценить человека по пяти самым основным критериям индивидуальности: открытость, добросовестность, экстраверсия, согласованность и нервозность. Компьютерные прогнозы основывались на

том, какие статьи, видео, музыка и другие элементы представляли интерес для пользователя в сети Facebook.

Методология исследования

Исследователи собрали самооценки 86,220 волонтеров, используя длинную анкету со 100 вопросами для выявления индивидуальных черт. Близкие люди волонтеров, включая друзей в сети Facebook и членов семьи, выражали свое суждение об индивидуальности их характера, используя анкетный опрос с 10 элементами. Также была получена компьютерная оценка индивидуальности на основе их «лайков» в Facebook.[9]

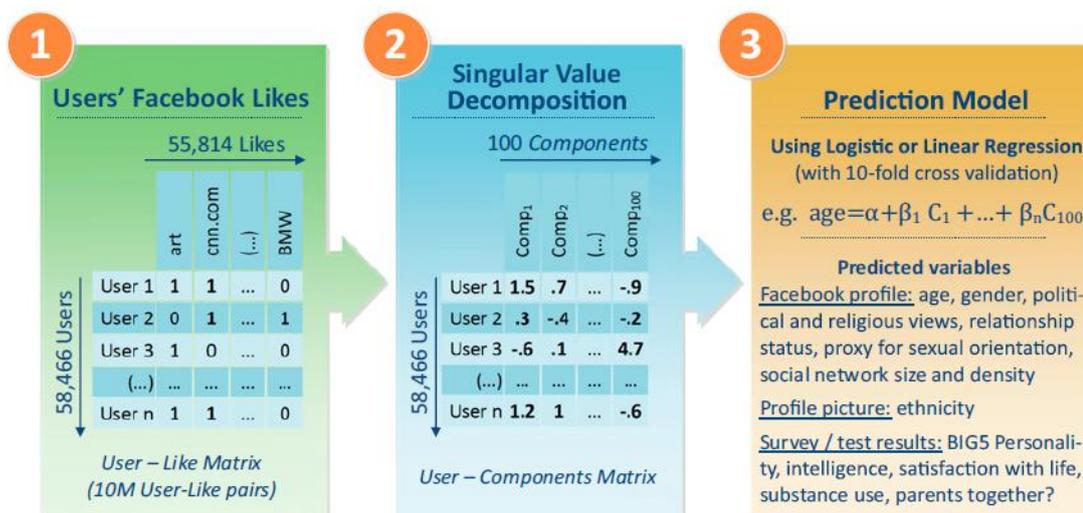


Рисунок 1 – Модель исследования профилей волонтеров в Facebook

Далее проводилось исследование, основанное на данных 58,466 американских волонтеров, полученных с помощью приложения myPersonality, которое находится в Facebook (www.mypersonality.org/wiki). Аккаунты включали информацию о профиле в Facebook, список «лайков» ($n = 170$ «лайков» на человека в среднем), психометрические экзаменационные отметки и информация из анкеты. Пользователи и их «лайкам» были представлены в виде матрицы «Пользователь - «лайк»», была поставлена 1, если была ассоциация между пользователем и его «лайком» и 0, если нет. Размерность матрицы «Пользователь - «лайк»» была уменьшена за счет использования сингулярного разложения (SVD). Числовые переменные, такие как возраст или интеллект были предсказаны, с помощью линейной модели регресса, тогда как дихотомические переменные, такие как пол или сексуальная ориентация были предсказаны с помощью логистической регрессии. В обоих случаях, была использована 10-кратная перекрестная проверка и $k = 100$ основных SVD компонентов. Для сексуальной ориентации, семейного положения родителей и потребления лекарств использовались только $k = 30$ главных компонентов SVD из-за меньшего числа пользователей, у которых была открыта эта информация [5].

Модель исследования представлена на рисунке 1. Были выбраны черты и признаки, которые показывают, насколько точным может быть такой

анализ, включающий: “сексуальную ориентацию”, “этническое происхождение”, “политические взгляды”, “религию”, “индивидуальность”, “интеллект”, “удовлетворение жизнью” (SWL), вредные привычки (“алкоголь”, “наркотики”, “сигареты”), “оставались ли родители человека в законном браке до достижения им 21 года”, и основные демографические признаки, такие как: “возраст”, “пол”, “семейное положение”, и “размер и плотность сети дружбы”.

Были установлены пять факторов модели оценки индивидуальности ($n = 54,373$) с помощью анкетного опроса International Personality Item Pool (IPIP) с 20 пунктами. Уровень интеллекта ($n = 1,350$) был измерен с помощью Raven’s Standard Progressive Matrices (SPM) и SWL ($n = 2,340$) был измерен с помощью шкалы SWL. Возраст ($n = 52,700$; среднее число, $\mu = 25.6$; $SD = 10$), пол ($n = 57,505$; 62% женщин), семейное положение (“один” / “в отношениях”; $n = 46,027$; 49% один), политические взгляды (“либеральный” / “консерватор”; $n = 9,752$; 65% либерал), религия (“мусульманин” / “христианин”; $n = 18,833$; 90% христианин), и социальная сетевая информация Facebook [$n = 17,601$; средний размер, $\sim X = 204$; диапазон межквартиля (IQR), 206; средняя плотность, $\sim X = 0.03$; IQR, 0.03] были получены из профилей пользователей Facebook. Потребление пользователями алкоголя ($n = 1,196$; 50% потребляет), наркотиков ($n = 856$; 21% принимает наркотики), и сигареты ($n = 1211$; 30% курит) и остались ли родители пользователя вместе, до достижения пользователем 21 года ($n = 766$; 56% остались вместе), были получены, используя интернет-опросы. Визуальный осмотр фотографий профиля использовался для определения этнического происхождения пользователей ($n = 7,000$; 73% белый; 14% афроамериканец; 13% другой). Сексуальная ориентация была определена, исходя из профиля в Facebook в области, «представляющей интерес». Пользователи, заинтересованные только представителями своего пола, были выделены в группу как гомосексуалисты (мужчины - 4.3%; женщины - 2.4%), тогда как пользователи, заинтересованные лицами противоположного пола, были отмечены как гетеросексуальные [9].

Прогнозирование результатов Дихотомических Переменных

Точность предсказания дихотомических переменных показана на рисунке 2 Она выражена в значении площади под кривой (AUC), который эквивалентен вероятности правильной классификации двух беспорядочно отобранных пользователей, по одному от каждого класса (например, мужчина и женщина). Самая высокая точность была достигнута для критерия этнического происхождения и пола. Афроамериканцы и евроамериканцы были правильно классифицированы в 95% случаев, мужчины и женщины были правильно классифицированы в 93% случаев. Христиане и мусульмане были правильно классифицированы в 82% случаев, и похожие результаты были достигнуты для демократов и республиканцев (85%). Сексуальную ориентацию было легче определить у мужчин (88%), чем у женщин (75%).

Хорошая точность предсказания была достигнута для семейного положения и употребления веществ (между 65% и 73%).

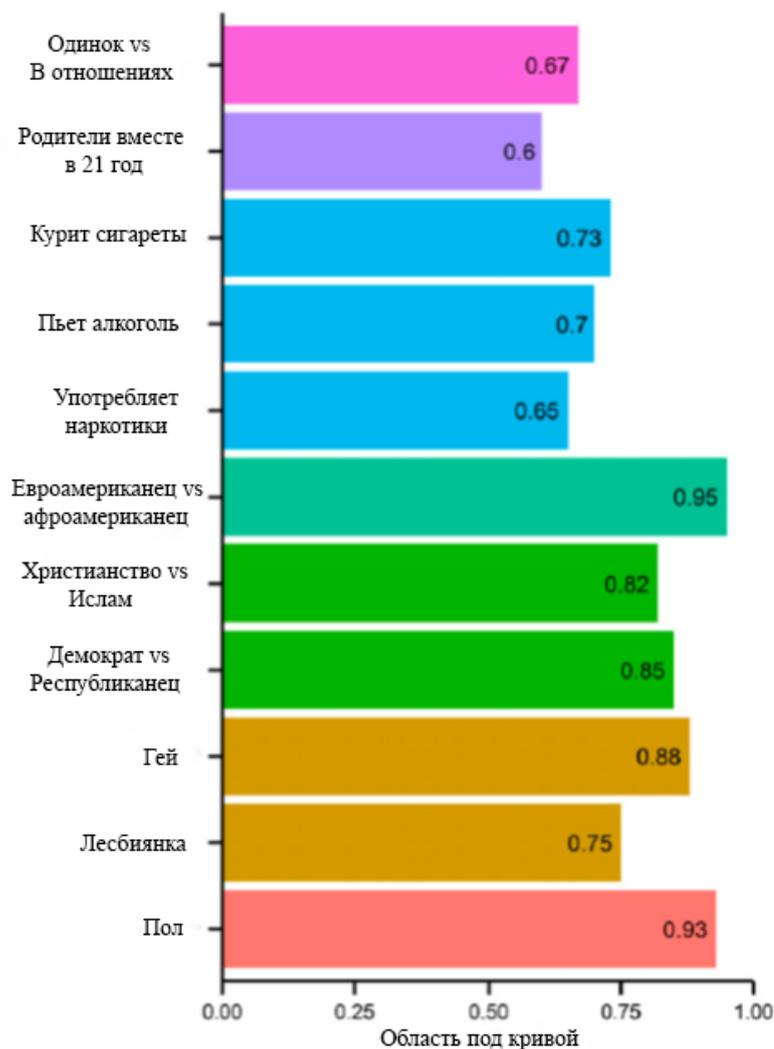


Рисунок 2 – Точность предсказания классификации для дихотомических/деливших на две части признаков выражена AUC

Относительно низкая точность для семейного положения может быть объяснена ее временной изменчивостью по сравнению с другими дихотомическими переменными (например, пол или сексуальная ориентация). Точность модели была самой низкой (60%), при определении, остались ли родители пользователя вместе или развелись до достижения им 21 года. Хотя известно, что развод родителей действительно имеет долгосрочное влияние на благополучие молодых людей. Примечательно, что это можно увидеть через их «лайки» в Facebook. У пользователей, родители которых развелись, есть более высокая вероятность проявления симпатии к утверждениям, связанным с отношениями. Например: «Если я с тобой, значит, я только с тобой, и никто другой мне не нужен» [6].

Предсказание числовых переменных

Точность предсказания числовых переменных, как выражено коэффициентом корреляции Пирсона – коэффициент корреляции момента между фактической и предсказанной ценностью, показана на рисунке 3. Самая высокая корреляция была получена для возраста ($r = 0.75$), сопровождаемой плотностью ($r = 0.52$) и размером ($r = 0.47$) сети дружбы Facebook. Были очень близко с этим черты индивидуальности из “Открытости” ($r = 0.43$), “Экстраверсия” ($r = 0.40$), и “Интеллекта” ($r = 0.39$). Остающиеся черты индивидуальности и SWL были предсказаны с несколько более низкой точностью ($r = 0.17$ к 0.30). Психологические черты - примеры скрытых черт (т.е., черты, которые не могут быть непосредственно измерены). Как следствие их ценности могут только приблизительно измерены, например, при оценке ответов на вопросы анкеты [7].



Рисунок 3 – Точность предсказания числовых переменных

Точность предсказания регресса для числовых признаков и черт выраженная коэффициентом корреляции Пирсона между предсказанным и фактическим значением атрибута; все корреляции значительны в $P < 0.001$ уровня. Диаграмма указывает на точность анкетного опроса.

Доступный объем данных и точность прогноза

Предоставленные на сегодняшний день результаты основаны на людях, количество «лайков» которых было от 1 до 700. Среднее число «лайков» на человека составляло 68 (IQR, 152). Используя подвыборку ($n = 500$) пользователей, у которых по крайней мере было 300 «лайков», были построены прогнозирующие модели на основе беспорядочно отобранного подмножества $n = 1, 2, \dots, 300$ «лайков». Результаты представлены на рисунке 4 [9].

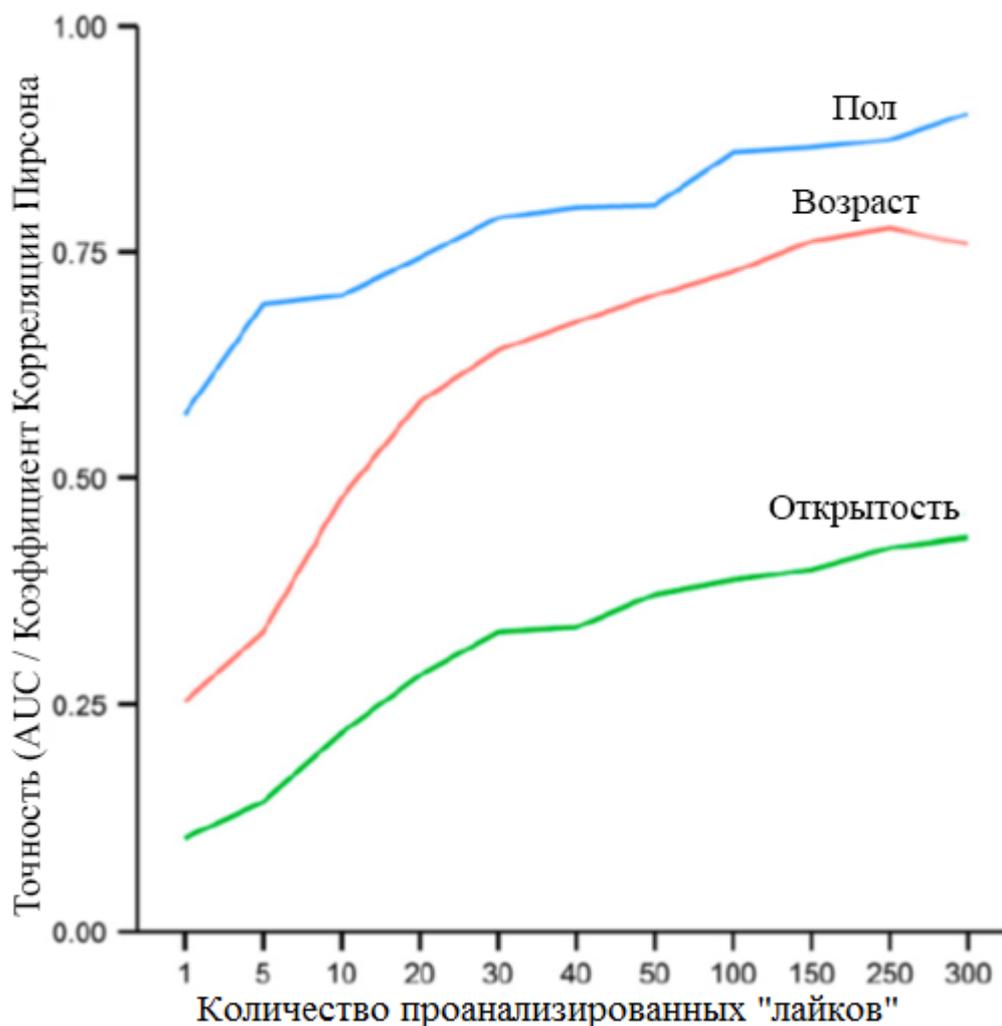


Рисунок 4 – Оценка точности прогноза

Точность отобранных предсказаний как функция количества доступных «лайков». Точность выражена как AUC (пол) и коэффициент корреляции Пирсона (возраст и Открытость). Приблизительно 50% пользователей в этом образце имели по крайней мере 100 «лайков», и приблизительно 20% имели, по крайней мере, 250 «лайков» [8].

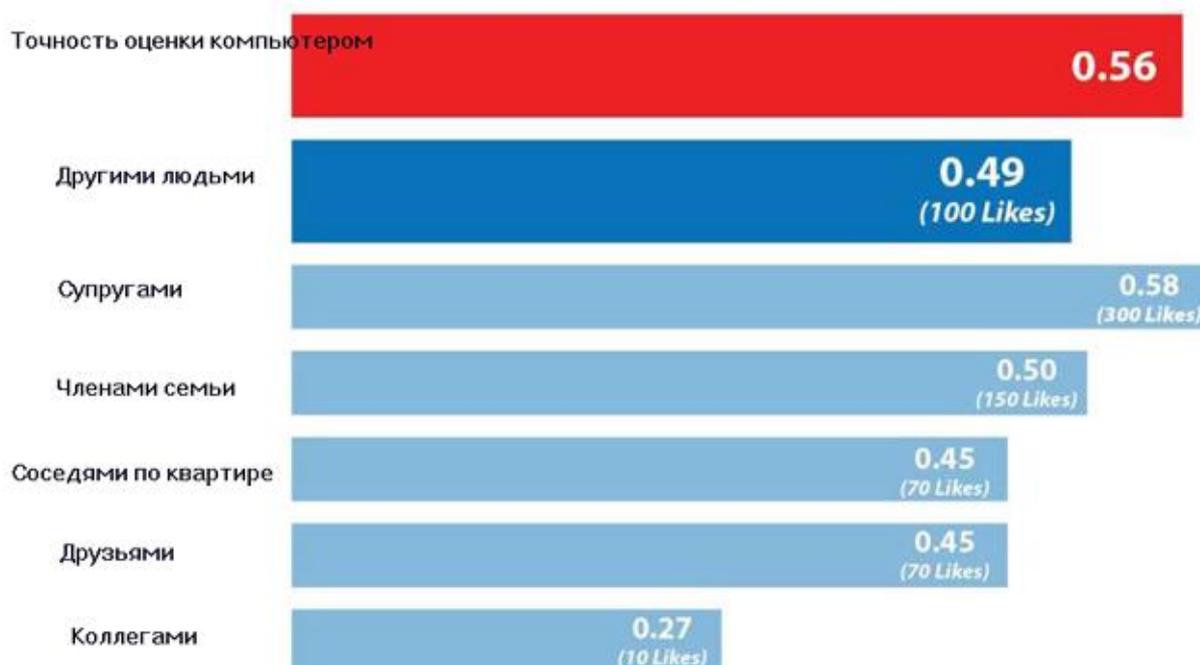


Рисунок 5 – Сравнение точности самооценки человека с оценкой компьютера и другими людьми

Результаты показали, что компьютер мог более точно предсказать индивидуальность характера, чем коллега с работы, анализируя всего 10 «лайков»; больше, чем друг или сосед по комнате с 70, член семьи со 150, и супруга с 300 «лайков» (рисунок 5).

Учитывая, что среднестатистический пользователь Facebook имеет приблизительно 227 лайков (и это число постоянно растет), у искусственного интеллекта есть потенциал, чтобы знать нас лучше, чем самые близкие для нас люди [2].

Почему машины лучше оценивают индивидуальность, чем люди?

Бытует мнение, что у компьютеров есть несколько главных преимуществ по отношению к людям в области анализа индивидуальности. Прежде всего, они могут сохранять и получать большие объемы информации и анализировать все эти данные с помощью алгоритмов.

По мнению исследователей, это обеспечивает точность, так как человеческому разуму трудно дать оценку одному или двум примерам. Также сложность представляет человеческая тенденция использовать слишком нерациональные способы мышления.

Тем не менее, исследователи признают, что определение некоторых черт индивидуальности лучше всего определяются людьми [7].

Выводы

Таким образом, в ходе работы была рассмотрена система построения скоринг-моделей оценки кредитоспособности клиентов по их цифровым следам. Располагая данными по кредитной истории своих заемщиков, каждый банк сможет на основе предложенной системы разработать свою

собственную скоринг-модель для оценки кредитоспособности потенциальных заемщиков, что позволит повысить обоснованность принятия кредитных решений и снизить уровень кредитных рисков.

Из этого следует, что широкий спектр личных качеств людей, начиная от сексуальной ориентации и заканчивая интеллектом, может быть автоматически и точно определен с использованием их «лайков» в Facebook. Сходство между «лайками» в Facebook и другими широко распространенными видами «цифровых следов», такие как история просмотров, поисковые запросы или история покупок показывают, что потенциал выявления данных пользователей вряд ли ограничится одними «лайками». Кроме того, широкое разнообразие данных, представленное в этом исследовании, показывает, что, учитывая соответствующие учебные данные, возможно, будет выявлено и другие закономерности.

Предсказание индивидуальных данных и предпочтений пользователей может быть использовано для оценки психоэмоциональных качеств пользователей. Кроме того, актуальность маркетинга и рекомендации продуктов и услуг могут быть улучшены путем добавления психологических данных для моделей текущего пользователя.

Более того, цифровые записи поведения могут обеспечить удобный и надежный способ измерения психологических особенностей. Автоматизированная оценка на основе больших выборок поведения может быть не только более точным и менее подверженным обману и искажению, но также может дать оценку для определения тенденций.

Библиографический список

1. Приложение на базе социальной сети Facebook, осуществляющее сбор и анализ персональных данных (с согласия пользователей) - myPersonality Project // <http://mypersonality.org> URL: <http://mypersonality.org/wiki/doku.php> (дата обращения: 22.04.2018).
2. Малюгин В. И. Исследование эффективности алгоритмов классификации заемщиков банков на основе балансовых коэффициентов / В.И. Малюгин, О.И. Корчагин, Н.В. Гринь // Банковский Вестник. 2015. №7. С. 26–33.
3. Connelly V. S., Ones D. S. An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity // Psychological bulletin. 2010. Т. 136. №. 6. С. 1092.
4. Funder D. C. Accurate personality judgment // Current Directions in Psychological Science. 2012. Т. 21. №. 3. С. 177-182.
5. Kahn Jr P. H. et al. Do people hold a humanoid robot morally accountable for the harm it causes? // Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction. ACM, 2012. С. 33-40.
6. Kahn Jr P. H. et al. Psychological intimacy with robots?: using interaction patterns to uncover depth of relation // Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction. IEEE Press, 2010. С. 123-124.

7. Kenny D. A. A general model of consensus and accuracy in interpersonal perception //Psychological Review. 1991. T. 98. №. 2. C. 155.
8. Macrae C., Quadflieg S. Perceiving people. Handbook of Social Psychology. New York: McGraw-Hill
9. Kosinski M., Stillwell D., Graepel T. Private traits and attributes are predictable from digital records of human behavior //Proceedings of the National Academy of Sciences. 2013. T. 110. №. 15. C. 5802-5805.