

Применение социальных сетей для анализа и прогнозирования

Гальчук Ю.В.

Актуальность темы

Анализ данных, полученных из социальных сетей, с каждым днём становится популярнее. Социальные сети являются уникальным источником данных о личной жизни и интересах обычных людей. Это открывает невероятные возможности для решения исследовательских и бизнес-задач (из-за недостатка данных раньше подобные задачи невозможно было решить), а также создания дополнительных сервисов и приложений для пользователей социальных сетей. Кроме того, этим предопределяется повышенный интерес к сбору и анализу социальных данных со стороны компаний и исследовательских центров [1].

Аналитическое агентство Gartner в 2012 году опубликовало отчёт под названием "Цикл ажиотажа для развивающихся технологий". Согласно отчёту, технологии "Социальная аналитика" и "Большие данные" в настоящее время находятся на т.н. "пике завышенных ожиданий". В частности, исследованиями социальных данных активно занимаются университеты Карнеги-Меллон, Стэнфорд, Оксфорд, INRIA, а также компании Facebook, Google, Yahoo!, LinkedIn и многие другие. Компании-владельцы сервисов онлайн-социальных сетей (Facebook, Twitter) активно инвестируют в разработку усовершенствованных инфраструктурных (Cassandra, Presto, FlockDB, Thrift) и алгоритмических (новые алгоритмы поиска и рекомендации пользователей, товаров и услуг) решений для обработки больших массивов пользовательских данных. Возникают и успешно развиваются коммерческие компании, предоставляющие услуги по доступу к хранилищам социальных данных (GNIP), сбору социальных данных по заданным сценариям (80legs), социальной аналитике (DataSift), а также расширению существующих платформ с помощью социальных данных (FlipTop) [2].

Таким образом, специалисты из исследовательских центров и компаний по всему миру используют данные социальных сетей для моделирования социальных, экономических, политических и других процессов от персонального до государственного уровня с целью разработки механизмов воздействия на эти процессы, а также создания инновационных аналитических и бизнес-приложений и сервисов.

В данной статье будут рассмотрены различные направления применения анализа социальных сетей.

Степень изученности вопроса

Работа П. Холланда [3] положила начало применению эмпирических статистических сетевых моделей для анализа социальных сетей (p^* , или экспоненциальные модели случайных графов, Exponential Random Graph Models, ERGM). Впоследствии идеи Холланда были развиты С. Вассерманом, Т. Снайдерсом, Ф. Паттинсон, Г. Робинсом и др. [4]. В рамках этой модели наблюдаемая социальная сеть рассматривается как одна из возможных реализаций сетей со схожими эндогенными и экзогенными параметрами. Задача моделирования заключается в выявлении структурных и иных характеристик социальной системы, отличающих исследуемую сеть от случайных.

Воронкин Алексей Сергеевич, аспирант кафедры государственной службы, администрирования и управления Государственного учреждения «Луганский национальный университет имени Тараса Шевченко», в своей работе «Социальные сети: эволюция, структура, анализ» провел детальный обзор истории развития социальных сетей и их анализа. [5]

Во второй половине XX в. существенно расширился системный анализ социальных сетей в работах таких исследователей, как Р. Соломонофф, С. Берковиц, С. Боргэтти, Р. Берт, К. Карли, К. Фост, Д. Нок, П. Марсден, Н. Маллинс, А. Рапопорт, С. Уоссермэн, Б. Веллмэн, Д.

Вайт, В. Харрисон и многих других. В 1950-е гг. работы английских антропологов Дж. Барнеса и Э. Ботт привлекли внимание к более сложным социальным феноменам, связанным с принадлежностью к социальным сетям. Сам термин «социальная сеть» был введен в 1954 году социологом Джеймсом Барнсом [6]. Из многих антропологических трудов того времени следует выделить работу С. Найдела «Теория социальной структуры» (1957 г.), в которой проведено четкое разделение «структуры» и «функции», что соответствует принципам современного сетевого анализа.

В 1959–1968 гг. венгерские математики Пол Эрдос и Альфред Реньи опубликовали серию статей, описывающих принципы формирования социальных сетей. Ученые впервые применили математическую теорию для иллюстрации принципа построения социальных сетей. Предложенная теория случайных графов позволила описывать сложные сети, не имеющие очевидных принципов построения [7].

В 1970-х гг. американский социолог Марк Грановеттер [8] выявил, что внутри социальных сетей «слабые» связи (например, наши соседи, знакомые, знакомые знакомых, формальные контакты на работе) имеют большее значение, чем «сильные» (к примеру, наши родственники и друзья). Объясняется это тем, что информация быстрее и шире распространяется именно посредством «слабых» связей [9]. По его мнению, «слабые» связи «совершенно необходимы для расширения возможностей взаимодействия пользователей и для их взаимодействия с сообществом, тогда как в результате «сильных» связей образуется локальная связь». На примере трудоустройства он показал, что с точки зрения поиска работы и прочих нужд в жизни связи с людьми, которых мы не очень хорошо знаем, более полезны. Это связано с тем, что через «сильные» связи люди делятся ограниченным объемом данных или ресурсов, «сильные» связи информационно избыточны, а, следовательно, они менее полезны друг другу. Таким образом,

преимущество в поиске работы имеет тот, кто имеет большую сеть пересекающихся контактов. Этот эффект М. Грановеттер назвал «силой слабых связей». Теоретическое обоснование тезиса о силе «слабых» связей предложил Рональд Бэрт в своей теории «структурных дыр» [10], под которыми понимается существование коммуникационно не связанных между собой пространств. Теория основана на идеях силы «слабых» связей, значимости посредничества в социальной сети, теории эксклюзивного обмена, а также на различении трех форм капитала – финансового, человеческого и социального. Наличие большого количества «слабых» связей действительно дает индивиду преимущества в получении информации [11]. Дункан Уоттс и Стивен Строгач развили теорию социальных сетей, и в числе многих других открытий ввели понятие коэффициента кластеризации – степени близости между неоднородными группами (например, когда человек расширяет сеть своих связей за счет людей, которых он лично не знает, но знают его друзья) [12].

С развитием сетевых технологий и появлением новых средств коммуникаций в 80–90-х гг. употребление термина «социальные сети» постепенно смещается из социологии в сферу информатики. В 1984 году Национальный фонд науки США основал обширную межуниверситетскую сеть NSFNet. Со второй половины 1980-х гг. реализуются первые прообразы современных социальных сетей – CompuServe, Prodigy и The Well. Зарождаются системы мгновенных сообщений – IRC и ICQ. В 1995 году появляется американский портал Classmates. Начало 2000-х гг. принято считать расцветом эпохи социальных Интернет-сетей – в 2003–2004 гг. запущены сети LinkedIn, MySpace и Facebook. И если LinkedIn создавалась с целью установления и поддержания деловых контактов, то владельцы MySpace и Facebook сделали ставку в первую очередь на удовлетворение человеческой потребности в самовыражении. В 2006 г. открылся проект Twitter, а также

популярная на территории стран СНГ социальная сеть «ВКонтакте», имеющая на сегодняшний день самый большой видео- и аудио-хостинг.

Сообщения в сети и фондовый рынок

ITI Capital (Инвестиционная Компания «Ай Ти Инвест») в своём блоге[13] на популярном среди разработчиков сайте habrahabr.ru рассказывала, что одним из наиболее перспективных и популярных направлений деятельности исследований является прогнозирование ситуации на фондовом рынке на основе различной информации. Для этого, в том числе, применяются и данные о тональности сообщений, опубликованных в интернете (sentiment analysis). В данной статье был получен на вопрос возможности создания эффективной торговой стратегии с помощью данного метода.

Сообщения в социальных сетях могут оказывать серьезное влияние на ситуацию на фондовом рынке. Широко обсуждаются случаи, когда некоторые трейдеры зарабатывают значительные суммы буквально с помощью одного «твита»[14]. Иногда этим даже пользуются злоумышленники, которые, к примеру, создают поддельные аккаунты аналитических компаний, публикуя в них сообщения, которые могут повысить или обвалить цену акций конкретной компании.

В свою очередь, разработчики торговых систем стараются создать алгоритмы, которые могли бы генерировать рекомендации о совершении покупки или продажи на основе доступных в сети данных. Существует множество разновидностей подобных алгоритмов. Некоторые из них подразумевают измерение интереса к теме финансов, который может свидетельствовать о скорых движениях рынка — для этого могут использовать, к примеру, сервис Google Trends.

Кроме того, разрабатываются и системы, которые анализируют тональность публикуемых новостей или обзоров финансовых экспертов.

Профессор Венского университета экономики и бизнеса Рональд Хохрайтер (Ronald Hochreiter) опубликовал [15] описание эксперимента по созданию торговой стратегии, которая использует для создания прогнозов оценку тональности сообщений в социальных сетях и сообществах.

В ходе эксперимента Хохрайтер использовал исторические данные о ценах акций, входящих в индекс Доу-Джонса в период с 2010 до конца 2013 года. При этом для сравнения результатов работы стратегии, основанной на анализе тональности сообщений в сети, использовался классический подход к формированию портфолио финансовых инструментов на основе анализа стандартного отклонения финансовых результатов тех или иных акций на исторических данных.

Стратегия, основанная на анализе тональности сообщений, показала лучшие характеристики рискованности операций, также при ее использовании была зафиксирована меньшая максимальная просадка портфеля.

В свою очередь представители ИТ-департамента из нью-йоркского Университета Stony Brook Венбин Жанг (Wenbin Zhang) и Стивен Скиена (Steven Skiena) в своей работе проанализировали [16] связь между тональностью сообщений в социальных сетях и публикаций в медиа и реальными результатами конкретных акций. Для этого они загрузили исторические данные о 3238 акций в период с 2005 по 2009 год. Вот, что им удалось выяснить:

Исследователи обнаружили связь между числом публикаций и количеством обсуждений, а также объёмом торгов — чем более популярной оказывалась компания, тем больше сделок совершалось с ее акциями (хотя сила этой связи зависит от сектора бизнеса — авиакомпании крайне чувствительны к тональности медиа, ИТ-компании — в меньшей степени).

Влияние социальных сетей (например, Twitter) носит более отложенный эффект, по сравнению с сообщениями в медиа — иногда результат обширных обсуждений в Twitter отражался на ценах акций только на следующий торговый день или еще днем спустя, в то время как публикации в крупных СМИ немедленно способствовали росту или падению котировок.

Стратегии, основанные на анализе данных СМИ, блогов или социальных сетей показывают лучшие результаты на коротких интервалах — рынок обычно реагирует на новости довольно быстро, поэтому более длинный период удержания позиции ничего не дает трейдеру(рис.1).

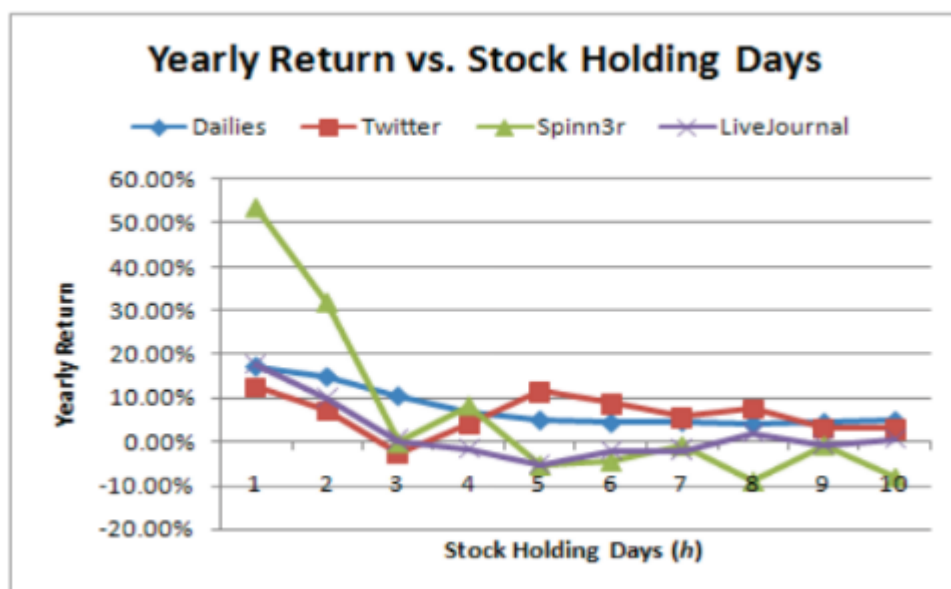


Рис. 1 Автокорреляция Lag-k-Day полярности для Dailies, Twitter, Spinn3r и LiveJournal соответственно [16]

Кроме того, нет смысла и выбирать на основе анализа тональности большое количество акций — чем больше выбранных таким образом инструментов в портфолио, тем ниже его общие результаты(рис.2):

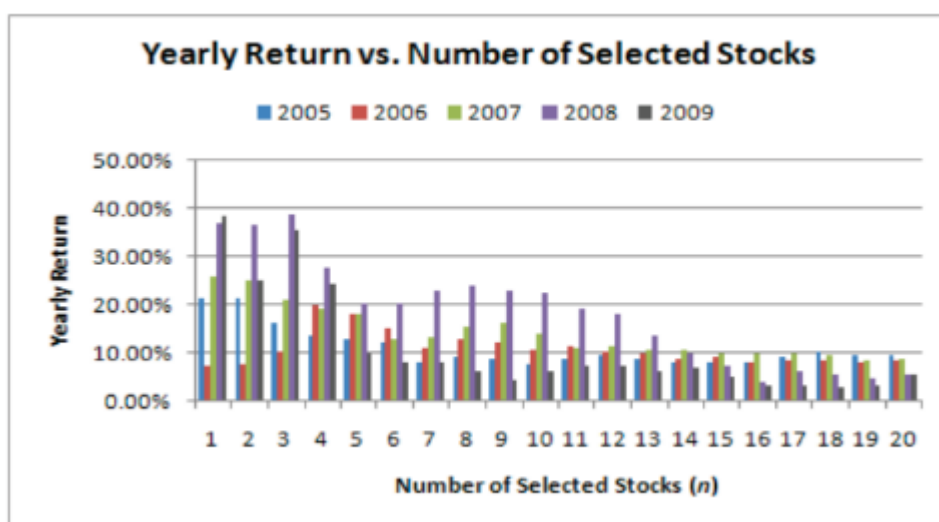


Рис.2 Годовой доход против количества выбранных верхнего и нижнего запасов для депозитария Dailies. (n от 1 до 20) [16].

Исследования показывают, что между тональностью сообщений в социальных сетях, блогах и публикациях в медиа и положением дел на финансовых рынках действительно существует определенная связь.

Ученым удастся разрабатывать стратегии, которые показывают неплохие результаты на исторических данных.

Веб-аналитика для повышения эффективности сайта кафедры высшего учебного заведения

Цеслив Ольга Владимировна, кандидат технических наук, доцент, кафедра математического моделирования экономических систем, Национальный технический университет Украины «Киевский политехнический институт», в своей работе «Веб-аналитика для повышения эффективности сайта кафедры высшего учебного заведения»[17] исследовала эффективность сайта кафедры с помощью Google Analytics, с целью улучшения его конкурентоспособности. В результате исследования проанализированы возможности Google Analytics, который позволяет: получить эмпирические данные о посетителях сайта кафедры; сформировать количественные отчеты: размер аудитории, количество посетителей за различные промежутки времени, количество просмотров, количество возвратов, размер постоянной аудитории; сформировать качественные отчеты: характеристика и сегментация аудитории, источники, с каких сайтов, поисковых систем пришли пользователи. Исследована возможность применения веб-аналитики для улучшения конкурентоспособности кафедры во время вступительной кампании. Выделены новые направления маркетинговой деятельности: использование рекламных акций в социальных сетях, обновление данных на страницах с большим показателем отказов, доработка контента страниц, на которых пользователи не останавливаются.

Для кафедры важно не только привлечь посетителей, но и сделать их своими абитуриентами. В работе был составлен портрет аудитории: распределение посетителей по регионам (рис. 3). Основная часть киевляне - 76%, 17,95% - посетители разных областей Украины.



Рис. 3. Посещение сайта из различных регионов Украины[17]

Углубившись в изучение действий пользователей на сайте, автор проанализировала, сколько секунд посетители были на сайте и сколько страниц просмотрели (рис. 4).

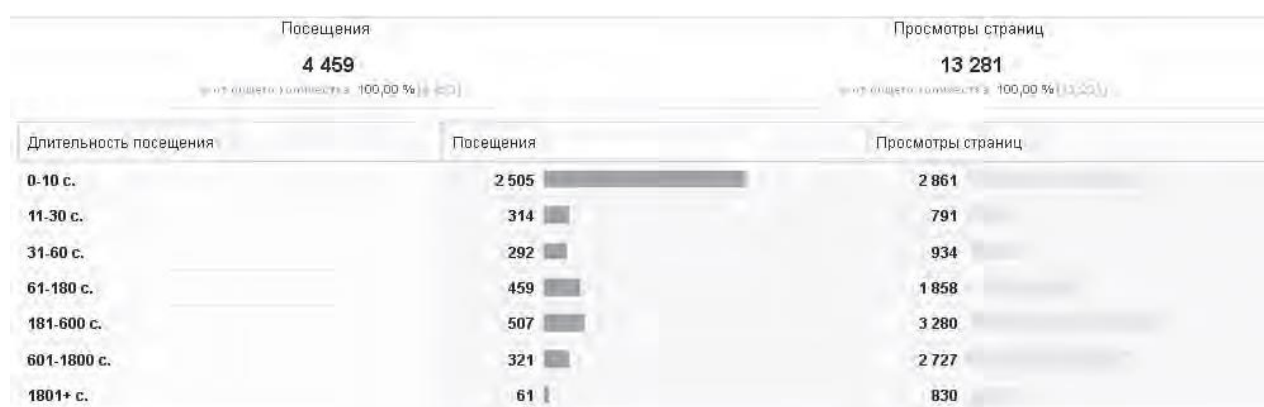


Рис. 4. Количество просмотренных страниц и время пребывания на сайте[17]

Далее автор перешла к содержанию сайта, чтобы выяснить, какие страницы пользуются наибольшей популярностью.

Анализ этих данных (рис.5) показал, что наибольшей популярностью пользуются страницы с информацией о преподавателях. Для первой части аудитории (это студенты) важная информация о программах экзаменов и консультации. Вторая часть - абитуриенты, которые интересуются условиями поступления в вуз.

☐	Сторінка	Перегляди сторінок ? ↓	Унікальні перегляди сторінок ?
		13 281 % від загальної кількості: 100,00% (13 281)	10 547 % від загальної кількості: 100,00% (10 547)
☐	1. / 	4 335	3 311
☐	2. /index.php?page=spirv 	957	799
☐	3. /index.php?page=studentam 	448	340
☐	4. /index.php?page=g 	438	332
☐	5. /index.php?page=magistrprog 	424	325
☐	6. /index.php?page=3kurs 	328	200
☐	7. /index.php?page=prog 	296	232
☐	8. /index.php?page=abitur 	277	208
☐	9. /index.php?page=kontakt 	255	199
☐	10. /index.php?page=cons 	218	170

Рис. 5. Наиболее популярные страницы[17]

Заключение

В статье было рассмотрено применение социальных сетей для анализа фондового рынка. Исследования показывают, что между тональностью сообщений в социальных сетях, блогах и публикациях в медиа и положением дел на финансовых рынках действительно существует определенная связь. Ученым удастся разрабатывать стратегии, которые показывают неплохие результаты на исторических данных.

Элементы подобных стратегий уже сейчас используют в своей практике некоторые финансовые компании и хедж-фонды. Однако полностью полагаться на социальные сети пока что рискованно.

Вторым исследуемым вопросом был анализ данных для повышения популярности кафедры учебного заведения среди потенциальных абитуриентов. Анализ показывает, что мало используются рекламные акции в социальных сетях, это требует доработки. Высокий уровень показателя отказов говорит о случайных

посетителей и о необходимости обновлении данных. Контент страниц, на которых пользователи не останавливаются, нуждается в доработке.

Можно сделать вывод, что анализ социальных сетей успешно применим как для продвижения товаров и услуг через Интернет, так и для прогнозирования различных социальных, экономических и других явлений.

Литература:

1. Антонова О. С., Колосюк А. А. Роль социальных сетей в системе образования // Финансовая система Украины: проблемы и перспективы развития в условиях трансформации социально-экономических отношений: материалы междунар. научно-практ. конф. (Севастополь, 16-18 мая 2013). – Севастополь: ДИАП, 2013. – С. 211–213.)
2. Key Trends to Watch in Gartner 2012 Emerging Technologies [Электронный ресурс]. – Режим доступа: <http://www.forbes.com/sites/gartnergroup/2012/09/18/key-trends-to-watch-in-gartner-2012-emerging-technologies-hype-cycle-2/>
3. Holland P. W., Leinhardt S. (1981) An Exponential Family of Probability Distributions for Directed Graphs // Journal of the American Statistical Association. Vol. 76. No 373. P. 33–50.]
4. Robins G. (2013) A Tutorial on Methods for the Modeling and Analysis of Social Network Data // Journal of Mathematical Psychology. Vol. 57. No 6. P. 261–274
5. Воронкин Алексей Сергеевич Социальные сети: эволюция, структура, анализ // ОТО. 2014. №1. [Электронный ресурс]. – Режим доступа: <http://cyberleninka.ru/article/n/sotsialnye-seti-evolyutsiya-struktura-analiz>
6. Barnes J. A. Class and committees in a Norwegian Island Parish // Human Relations. – 1954. V. 7. – P. 39–58. [Электронный ресурс]. – Режим доступа: <http://pierremerckle.fr/wpcontent/uploads/2012/03/Barnes.pdf>
7. Erdos P., Renyi A. On the evolution of random graphs // Publ. Math. Inst. Hungar. Acad. Sci. – 1960. V. 5. – P. 17–61
8. Granovetter M. The Strength of Weak Ties // American Journal of Sociology. – 1973. V. 78. – № 6. – P. 1360–1380
9. Стрельников А. Н. Социальные сети: механизмы работы и пути развития. Режим доступа: <http://www.rae.ru/forum2011/153/1796>

10. Burt R. Structural holes: the social structure of competition. – Harvard University Press, 1995. – 315 p
11. Жуликов С. Е., Жуликова О. В. Современные подходы к анализу социальных сетей // Гаудеамус: психолого-педагогический журнал. – 2012. – №2 (20). – С. 200–202
12. Watts D. J., Strogatz S. H. Collective dynamics of «small-world» networks // Nature. – 1998, June. Vol. 393. – P. 440–442
13. Можно ли создать алгоритм для торговли на бирже с помощью анализа тональности сообщений в интернете [Электронный ресурс]. – Режим доступа: <https://habrahabr.ru/company/itinvest/blog/282403/>
14. Соцсети, финансы и роботы: Как трейдер заработал \$2,4 млн за 28 минут с помощью одного твита [Электронный ресурс]. – Режим доступа: <https://geektimes.ru/company/itinvest/blog/268890/>
15. Computing trading strategies based on financial sentiment data using evolutionary optimization Ronald Hochreiter April 2015 [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1504.02972.pdf>
16. Trading Strategies to Exploit Blog and News Sentiment Wenbin Zhang and Steven Skiena [Электронный ресурс]. – Режим доступа: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1529/1904>
17. Веб-аналітика для підвищення ефективності сайту кафедри вищого навчального закладу / О. В. Цеслів // Бізнес Інформ. - 2013. - № 10. - С. 161-167. [Электронный ресурс]. – Режим доступа: http://nbuv.gov.ua/UJRN/binf_2013_10_31

Гальчук Ю.В.

Применение социальных сетей для анализа и прогнозирования

В данной статье было рассмотрено развитие социальных сетей и их применение для прогнозирования, анализа данных. Был сделан исторический обзор, а также изучены некоторые научные работы, связанные с выбранной темой. Так, было изучено влияние сообщений в социальных сетях на ситуацию на фондовых рынках. Также было показано, как можно использовать средства Google Analytics для продвижения сайта учебного заведения среди потенциальных абитуриентов.

В последующих работах планируется проведения собственного исследования в данной сфере.

Ключевые слова: социальные сети, анализ, онлайн-трейдинг, прогнозирование, продвижение, веб-аналитика.

Гальчук Ю.В.

Застосування соціальних мереж для аналізу і прогнозування

У даній статті було розглянуто розвиток соціальних мереж і їх застосування для прогнозування, аналізу даних. Був зроблений історичний огляд, а також вивчені деякі наукові роботи, пов'язані з обраною темою. Так, було вивчено вплив повідомлень в соціальних мережах на ситуацію на фондових ринках. Також було показано, як можна використовувати засоби Google Analytics для просування сайту навчального закладу серед потенційних абітурієнтів.

У наступних роботах планується проведення власного дослідження в даній сфері.

Ключові слова: соціальні мережі, аналіз, онлайн-трейдинг, прогнозування, просування, веб-аналітика.

Информация об авторе:

Гальчук Юлия Витальевна — студентка 5-го курса Харьковского национального университета имени В.Н.Каразина экономического факультета специальности «Экономическая кибернетика»

Інформація про автора:

Гальчук Юлія Віталіївна - студентка 5-го курсу Харківського національного університету імені В.Н.Каразіна економічного факультету спеціальності «Економічна кібернетика»