Раздел
Информационные процессы
Section
Information Processes

УДК: 311, 336

JEL: C55

Анализ отношения пользователей интернета
к исламской финансовой системе

Алескерова Н. М.

Заведующая отделом, докторант «Научно-Исследовательского Института
Экономических Реформ Министерства Экономики Азербайджанской
Республики. г. Баку, ул. Гасанбека Зардаби 88, AZ 1011, Азербайджан
nigar.kazimova@gmail.com

Аннотация. В статье исследовано отношение пользователей интернета к исламской финансовой системе с использованием данных Google Trends и WorldData.AI. Используя статистические данные опроса Google Trends, было сопоставлено отношение к исламским финансам, исламской экономике и исламскому банковскому делу. Также, анализируя новостные тексты, содержащиеся в источнике информации WorldData.AI об исламских финансах за последние 12 месяцев, было установлено, что отношение общества к исламским финансам является позитивным.

Ключевые слова: исламская финансовая система, Big Data, метод анализа текста, модель LDA, Google Trends.

An Analysis of the Attitudes of Internet Users Towards
the Islamic Financial System

Alasgarova N.M.

“Institute for Scientific Research on Economic Reforms of Azerbaijan Republic
Ministry of Economy” Branch Manager, PhD. Baku, Hasan bay Zardabi str. 88, AZ
1011 Azerbaijan
nigar.kazimova@gmail.com

Abstract. In this article has been explored the attitude of Internet users toward the Islamic financial system using data from Google Trends and WorldData.AI. An attitudes towards Islamic finance, Islamic economics and Islamic banking have been compared using from Google Trends survey statistics. Also, in the WorldData.AI. source has been analyzed that the news texts on Islamic finance for the last 12 months and found that the public attitude towards Islamic finance was positive.

Keywords: islamic financial system, Big Data, text analysis method, model LDA, Google Trends.

DOI: 10.31432/1994-2443-2021-16-3-32-44

Цитирование публикации: Алескерова Н. М. Анализ отношения пользователей интернета к исламской финансовой системе // Информация и инновации. 2021, Т. 16, № 3. с. 32-44. DOI: 10.31432/1994-2443-2021-16-3-32-44

Citation: Alasgarova N.M. An analysis of the attitudes of internet users towards the islamic financial system // Information and Innovations 2021, T. 16, № 3. p. 32-44. DOI: 10.31432/1994-2443-2021-16-3-32-44

Введение

Известно, что в настоящее время интернет стал одной из основных потребностей нашей жизни. Сегодня многие исследователи, используя в своей работе информацию с интернет-ресурсов, в социальных сетях, на новостных сайтах, пытаются расчитать показатели, не являющиеся официальной статистикой, изучить отношение общества к объекту исследования. В настоящее время растет интерес многих исследователей к анализу больших данных (Big Data) и текстов, методам интеллектуального

анализа. Для экономистов эти неструктурированные материалы могут рассматриваться как дополнительный источник информации для более детального изучения различных экономических явлений и составления прогнозов.

В связи с этим автор рассматривает оценку исламской финансовой системы и тональность текстов, написанных в этом направлении, используя статистические данные опроса Google Trends и WorldData.AI, а не официальной статистики.

Обзор литературы

Рост интереса к анализу больших данных (Big Data) привел к увеличению количества исследовательских работ, проводимых в этой области. Большая часть исследовательских работ в этом

направлении посвящена моделям краткосрочной оценки различных экономических явлений. В частности, специалисты Центрального Банка России, Т.В. Батура, М.В. Чаринцева, Е.В. Ба-

лацкий, М.А. Юревич, Н.Н. Куницына, Ю.А. Метель, К. Яковлева, О.В. Митина, А.С. Евдокименко, И.Е. Воронина, В.А. Гончаров и др. исследователи [1-4] провели изыскания и в своих работах объясняют методы анализа больших данных и текстов, реализуемые алгоритмы, преимущества и недостатки этих методов и так далее. Авторы применили метод текстового анализа при оценке экономической активности, инфляционных ожиданий, удовлетворенности клиентов коммерческих банков, тональности текстов и других экономических вопросов.

В своей работе И.О. Голощапова [5] предложила методику построения высокочастотных индикаторов экономических ожиданий населения на основе анализа больших данных с применением алгоритма LDA (Latent Dirichlet Allocation), который является моделью тематического моделирования. Другой исследователь Бауэр в своем исследовании показал значительное влияние новостей на динамику инфляционных ожиданий населения.

С. Бейкер, Р. Ниман, Д. Ардиа, К. Блуто, К. Боудт, А. Торсруд, Н. Бейлк и др. Исследователи [6] также рассмотрели вопрос построения экономических показателей с использованием метода текстового анализа.

В то же время исследования показывают, что использование Google Trends в моделях прогнозирования улучшает результаты прогнозов. В своем исследовании о продаже автомобилей Choi и Varian использовали статистику опроса Google Trends и пришли к выводу, что модель авторегрессии пер-

вого порядка временных рядов Google Trends обладает лучшими прогностическими свойствами, чем обычная модель авторегрессии первого порядка. Google Trends также положительно влияет на улучшение прогнозов по другим экономическим вопросам (безработица, доверие потребителей, туризм и т. д. [7].

1. Оценивание посредством Google Trends

Google Trends предоставляет пользователям Google доступ к неотфильтрованным поисковым запросам. Благодаря этому сервису можно определить, насколько интересна пользователям какая-либо тема в мире или на конкретной территории, по каким направлениям ведется поиск, узнать текущую экономическую активность и настроение.

Известно, что Google ежедневно обрабатывает миллиарды поисковых запросов. Учитывая большое количество данных в Google Trends, для ускорения их обработки используется только их выборка. Этот алгоритм позволяет очень быстро узнать отношение общества к любому новому событию. Поэтому в последнее время информация Google Trends стала широко использоваться аналитиками при составлении краткосрочных прогнозов различных экономических показателей.

В Google Trends за временной интервал можно принять режим реального времени — то есть последние 7 дней и режим нереального времени — при условии, что он не учитывает последние 36 часов с 2004 года [8]. В статье в ка-

честве временного интервала автором были приняты последние 12 месяцев. Прежде всего, 3 ключевых слова были использованы для определения того,

что больше всего интересует пользователей интернета: исламская экономика, исламские финансы и исламские банки.

Рис. 1. Динамика поиска пользователей интернета по ключевым словам
Источник: Google Trends

Из графика видно, что интерес интернет-пользователей к исламской экономике низок по сравнению с исламскими финансами и исламским банковским делом. Можно сказать, что интерес к исламским финансам и исламскому банковскому делу выше среднего (средняя оценка: исламская экономика — 7, исламские фи-

нансы — 61, исламское банковское дело — 72).

Также было исследовано, что больше всего интересует пользователей по странам с помощью приложения Google Trends, и региональная популярность, связанная с рассматриваемыми ключевыми словами, отражена на следующих диаграммах:

Рис. 2. Диаграмма региональной популярности по исламским финансам
Источник: Google Trends

Рис. 3. Диаграмма региональной популярности по исламскому банковскому делу
 Источник: Google Trends

Рис. 4. Диаграмма региональной популярности по исламской экономике
 Источник: Google Trends

Рис. 5. Самые популярные ключевые слова по исламскому банковскому делу
 Источник: Google Trends

Рис. 6. Самые популярные ключевые слова по исламским финансам

Источник: Google Trends

Наиболее популярные ключевые слова об исламских финансах и исламском банковском деле в рассматриваемый временной интервал (10.05.2020-02.05.2021) представлены на диаграмме ниже. Следует отметить, что из-за слабого интереса интернет-пользователей к исламской экономике была изучена статистика опроса по двум другим ключевым словам. (Рис. 5)

Также используя приложения Google Trends был сопоставлен интерес пользователей интернета во всем мире к исламским и коммерческим банкам.

Как видно, интерес к коммерческим банкам преобладает с большой разницей. Одна из основных причин этого заключается в том, что многие люди все еще не знают об исламской банковской системе.

Рис. 7. Интерес пользователей интернета к исламским и коммерческим банкам

Источник: Google Trends

2. Оценивание тональности текста

Для оценки того, является ли отношение к исламским финансам негативным или позитивным, будет использоваться метод анализа текста. Метод анализа текста позволяет решать вопросы по классификации, кластеризации, выводу идей и фактов, построению экспертных и вопросно-ответных систем, поиску по ключевым словам и др. Основным источником информации при проведении таких исследований являются новости, социальные сети, выступления официальных лиц и другая информация [6]. Эта методика предусматривает регулярное обновление собранной информации. Поскольку эти новости являются не только источником информации, они также позволяют изучать отношение журналистов, общества к экономическим процессам и превращать их в количественный показатель.

В качестве источника информации для оценки отношения к исламским финансам на первоначальном этапе должны быть определены основные новостные сайты. Поскольку тексты новостей составляют основу данного исследования, крайне важно выбрать источники новостей, более полно отражающие текущую экономическую ситуацию, имеющие широкую аудиторию. Выбор новостного сайта можно осуществить путем анализа рейтингов источников СМИ. При выборе источника новостей необходимо обратить внимание на следующее:

- Соответствие этих новостей рассматриваемой теме;

- Наличие долгосрочных временных рядов в интернете (не менее 3-4 лет);

- Возможность легкого поиска текста новостей с сайта (простота *web-scraping*) [6].

Также в исследовательской работе целесообразнее использовать несколько источников информации для получения более всеобъемлющего результата. Автором в качестве источника информации был выбран WorldData.AI. Выбор WorldData.AI в качестве базы данных был сочтен целесообразным, так как он концентрирует в себе информацию многих информационных агентств.

В исследовательской работе в качестве временных рядов были взяты последние пять лет. Всего было просмотрено 350 статей. Источники, из которых взяты новостные статьи, выбранные в соответствии с ключевым словом «исламские финансы», отражены на диаграмме ниже. Как видно из диаграммы, новостям об исламских финансах уделяется больше места на новостных сайтах Reuters (21,4%) и CNA Finance (20%).

Следующий этап — отбор новостных статей в экономических СМИ, соответствующих объекту исследования. Целью этого отбора является получение статей, наиболее соответствующих рассматриваемой экономической области в исследовательской работе. И это определяется на основе выбранных ключевых слов и словосочетаний, относящихся к области исследования. Отбор новостей СМИ по кон-

Рис. 8. Количество новостей по источникам
 Источник: WorldData.AI

кретной экономической сфере основывается на следующих принципах:

➤ прежде всего, необходимо включить в выборку все статьи, в которых обсуждается объект исследования, то есть статьи, в которых говорится о ключевых словах, определяющих этот объект в экономической теории. Например, поскольку объектом исследования являются исламские финансы, ключевыми словами для выбора можно использовать “исламские финансы”, “беспроцентные финансы” и т. д.

➤ на следующем этапе используются альтернативные словосочетания, относящиеся к объекту исследования. Например, “исламское банковское дело”, “исламские финансовые институты”, “мудараба” и др. [5].

Также при отборе текстов новостей можно использовать современные алгоритмы анализа текста, а именно тематическое моделирование.

Тематическое моделирование позволяет выбирать ключевые слова в выбранных текстах, определяя сходство тем на основе анализа частоты встре-

чающихся слов. Применение таких алгоритмов может сыграть важную роль в снижении доли субъективной экспертной оценки при составлении правил отбора, повышении качества правил за счет включения в процесс отбора слов и словосочетаний, определенных данным методом. Однако у этого метода также есть и недостатки. Таким образом, этот метод анализа текста характеризуется нестойкостью по отношению к изменяющемуся набору данных.

Каждая новостная статья может охватить одну или несколько тем в разной степени. Тематическое моделирование также позволяет автоматически сортировать все собранные статьи по темам. Тематическое моделирование позволяет выбирать ключевые слова из выбранных текстов, определяющих сходство тем на основе анализа частоты встречающихся слов. Отбор текстов в этом методе происходит за счет оптимизации параметров распределения слов и словосочетаний в статьях. Тематическое моделирование является одним из классических подходов к машинной обработке и в основном используется при проведении исследований по “обработке естественного языка” (Natural Language Processing, NLP). Такой подход используется для автоматической работы с большими объемами текстовой информации. Другими словами, тематическое моделирование направлено на определение кластеров слов и сочетаний слов в определенном наборе данных.

Для проведения тематического моделирования тексты должны пройти следующие этапы обработки:

- Процесс фильтрации — это удаление “лишних” символов из текста. Говоря лишние символы, имеется в виду нелингвистические символы, знаки препинания, повторяющиеся пробелы и многое другое;

- Токенизация — этот этап подразумевает разделение всего текста на отдельные слова;

- Удаление стоп-слов — то есть удаление слов, не имеющих существенного значения в теме. Список классических стоп-слов для каждого языка обычно определяется и регулярно обновляется лингвистами и является общедоступным. Сюда относятся отдельные буквы, союзы, числа, восклицательные знаки, модальные слова и т.д;

- Стемматизация — это приведение слова к их первоначальной форме. Например, слова «экономики» и «кризисы» в результате стемматизации заменяются словами «экономика» и «кризис». Для этого также существует специальное программное обеспечение. Например, MyStem [5].

Вкратце можно сказать, что тематические модели можно разделить на две группы: алгебраические и вероятностные. К алгебраическим тематическим моделям относятся модель векторного пространства (Vector Space Model, VSM) и модели Латентного семантического анализа (Latent Semantic Analysis, LSA). Алгебраические модели позволяют найти взвешенность слов

и на их основе оценить сходство рассматриваемых текстов.

Среди вероятностных моделей широкое распространение получили модели Латентного Размещения Дирихле (Latent Dirichlet Allocation, LDA) и вероятностно- Латентного Семантического Анализа (Probabilistic Latent Semantic Analysis, PLSA). PLSA — это усовершенствованная форма модели LSA и позволяет более качественно определять темы. Однако основным недостатком этой модели является то, что с увеличением количества документов увеличивается и количество параметров, что усложняет работу модели [9].

Хотя алгоритм LDA также имеет некоторые недостатки, он устраняет недостатки PLSA и имеет следующие преимущества по сравнению с рекомендациями экспертов по определению тем:

- автоматическая реализация этого алгоритма позволяет применять его к большому набору документов. Этот вопрос может создать серьезные трудности при экспертной оценке;
- результатом алгоритма LDA является не только надежная классификация тем для текстовых данных, но также может быть применен для выявления полученного набора тем для любой коллекции документов, включая новые тексты;
- алгоритм LDA не требует от исследователя предварительного уточнения правил поиска или ключевых слов для определения тем или других категорий в тексте.
- Темы и их вероятностные взаимодействия с ключевыми словами

определяются алгоритмом поиска оптимальной статистической модели в заданном текстовом диалоговом окне.

В последнее время применение модели LDA получило широкое распространение во многих экономических исследованиях. Например, А. Хуан, А. Торсруд, К. Яковлева И. Голощапова, М. Андреев и др. в своих работах исследователи и использовали метод LDA при выборе тем, комментарии для оценки различных экономических показателей.

В целом, по результатам тематического моделирования осуществляется ранжирование полученных тем или подбор значимых для объекта исследования тем не только на основе вероятности возникновения каждой темы в разные периоды, но и с учетом экспертной оценки значимости тем в контексте конкретной области исследования.

Модель LDA — это трехуровневая иерархическая байесовская модель, в которой каждый документ моделируется как набор не наблюдаемых, скрытых тем. Согласно модели, каждое слово в тексте принадлежит неизвестному нам тексту. Каждая тема моделируется из заранее определенных вероятностей темы. Математически это можно описать следующим образом:

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta)$$

где,

α и β — заданные векторы, т. е. параметры распределения Дирихле;

$\theta \sim \text{Dir}(\alpha)$ распределение тем в каждом документе;

$z_n \sim \text{Dir}(\beta)$ — распределение слов по каждой теме:

W_n — соответствие слов в документе темам [9].

Количество потоков должно быть задано в качестве начального параметра для модели LDA.

Применяя модель LDA, составляется перечень тем, представленных в виде отдельных слов (униграмм), наиболее характерных для каждой рассматриваемой темы. На рисунке

ниже в качестве примера представлена униграмма одного из текстов. На основании проведенных исследований (Холодилин К. и Томас Т.) обязательно читать все приведенные слова, чтобы узнать, к какой теме относится униграмма. Таким образом, первые 10 слов содержат 30% информации по теме, что может создать у исследователя достаточное представление о теме. Например, если посмотреть на следующую униграмму, видно, что оно посвящено исламским финансам и последствиям пандемии.

Рис. 9. Униграммa одного из текстов
Источник: WorldData.AI

Таким образом, на следующем этапе была определена тональность 350 статей, полученных из источника WorldData.AI. Когда говорят о тональности текста, то здесь подразумевают 3 случая: положительный, отрицательный и нейтральный. Когда в тексте отсутствует эмоциональная составляющая, он считается нейтральным. Тональность колеблется между -1 и +1 [9].

WorldData.AI сам представляет тональность каждого текста в готовом виде. На приведенной ниже диаграмме тональность рассматриваемых статей об исламских финансах выражена в процентах. (Рис. 10)

Как видно из диаграммы, основанной на анализе WorldData.AI, в мире преобладает более позитивное отношение к исламским финансам.

Рис. 10. Тональность текстов

Источник: Разработано автором на основе данных WorldData.AI

Заключение

Иногда более детальное изучение различных экономических явлений и составление прогнозов, помимо официальной статистики, оказывает положительное влияние на отношение общества к данному объекту исследования. С этой целью, используя данные Google Trends и WorldData.AI, исламская финансовая система была оценена альтернативным путем, и было определено, что интерес к исламским финансам и исламской банковской системе в мире больше, чем интерес к исламской экономике.

И это, на наш взгляд, не случайно. Потому что принцип деятельности исламской финансовой системы, инструменты финансирования, предлагаемые исламскими банками, отказ от процентов и другие причины привлекают многих людей, особенно малых и средних предпринимателей. Проведенное исследование также показывает, что в целом отношение общества к исламской финансовой системе является позитивным.

ЛИТЕРАТУРА

1. Балацкий Е.В., Юревич М.А. "Измерение инфляционных ожиданий: традиционные и новаторские подходы" Санкт-Петербург, 2018.
2. Куницына Н.Н., Метель Ю.А. «Применение метода интеллектуального анализа в оценке уровня удовлетворенности клиентов коммерческих банков», Вестник СГСЭУ. 2018. № 2.
3. Митина О.В., Евдокименко А.С. «Методы анализа текста: Методологические основания и программная реализация» Вестник ЮУрГУ, № 40, 2010.
4. Google Trends.
5. Голощапова И.О. "Разработка методики построения высокочастотных индикаторов экономических ожиданий населения на основе больших данных (на примере инфляционных ожиданий" Диссертация. Москва, 2018.
6. Центральный банк Российской Федерации "Тексты экономических новостей: полезное дополнение к официальной статистике?" Москва, 2018.
7. URL: <https://support.google.com/trends/answer/4365533?hl=ru>.

8. Воронина И., Гончаров В. Анализ эмоциональной окраски сообщений в социальных сетях // Компьютерная лингвистика и обработка естественного языка, Вестник ВГУ, серия: Системный анализ и информационные технологии. — 2015. — № 4.

9. Яковлева Ксения "Оценка экономической активности на основе текстового анализа" Москва, 2018.

REFERENCES

1. Balackij E.V., YUrevich M.A. "Izmerenie inflyacionnyh ozhidaniy: tradicionnye novatorskie podhody" Sankt-Peterburg, 2018.

2. Kunicyna N.N., Metel' YU.A. «Primenenie metoda intellektual'nogo analiza v ocenke urovnya udovletvorennosti klientov kommercheskih bankov», Vestnik SGSEU. 2018. № 2.

3. Mitina O.V., Evdokimenko A.S. «Metody analiza teksta: Metodologicheskie osnovaniya i programmaya realizaciya» Vestnik YUUrGU, № 40, 2010.

4. Google Trends.

5. Goloshchapova I.O. "Razrabotka metodiki postroeniya vysokochastotnyh indikatorov ekonomicheskikh ozhidaniy naseleniya na osnove bol'shih dannyh (na primere inflyacionnyh ozhidaniy)" Dissertaciya. Moskva, 2018.

6. Central'nyj bank Rossijskoj Federacii "Teksty ekonomicheskikh novostej: poleznoe dopolnenie k oficial'noj statistike?" Moskva, 2018.

7. URL: <https://support.google.com/trends/answer/4365533?hl=ru>.

8. Voronina I., Goncharov V. Analiz emocional'noj okraski soobshchenij v social'nyh setyah // Komp'yuternaya lingvistika i obrabotka estestvennogo yazyka, Vestnik VGU, seriya: Sistemnyj analiz i informacionnye tekhnologii. — 2015. — №4.

9. Yakovleva Kseniya "Ocenka ekonomicheskoy aktivnosti na osnove tekstovogo analiza" Moskva, 2018.