

**Построение прогнозной модели электорального поведения на основе сетевых технологий**

**Климова Дарья Сергеевна**

*Студент (бакалавр)*

Международный университет природы, общества и человека «Дубна», Дубна, Россия

*E-mail: dariaklmv@gmail.com*

Исследования, основанные на данных из социальных сетей, получают всё большее распространение в самых различных социальных науках. Несмотря на неравномерность выборочной совокупности населения, вызванную непропорциональным развитием технологий, Online Big Data продолжают привлекать исследователей из самых различных социальных наук, отчасти, из-за своих беспредельных размеров, а также из-за своей потенциальной способности прогнозировать действия людей в реальном мире. Одной из самых очевидных исследовательских отраслей, в связи с этим, является политический прогноз. Прогностические способности социальных сетей были и остаются по сей день популярной темой на международном уровне, однако точность прогнозов на их основе до сих пор является предметом диспута и имеет противоречивые отзывы. В данной работе осуществляется анализ ранее проведённых исследований, с целью ответить на вопрос о том, насколько рациональным является прогнозирование результатов выборов при помощи данных из социальных сетей и смогут ли онлайн-методы прийти на смену классическим опросам.

Первоначально специалисты предоставляли необычайно оптимистичные оценки прогностической способности данных, полученных из твиттер сообщений пользователей, как инструмента для построения модели электорального поведения. Выбор пал именно на эту социальную сеть из-за возраста пользователей (около 30 лет) и быстрого темпа роста аудитории [6]. Некоторые исследователи обнаружили закономерность, состоящую в том, что процентный объем сообщений с упоминанием кандидата или партии, с высокой вероятностью отражал реальные проценты, полученные политиками в ходе выборов. Метод показал хорошие результаты в одной из крупнейших стран Латинской Америки - Бразилии. В 2010 году была спрогнозирована не просто победа Русеф Дилмы, но и ее процентный отрыв от конкурентов со средней абсолютной погрешностью 4,07% [8]. Однако, пожалуй наиболее точного прогнозирования электорального поведения, на основе вышеизложенного метода, удалось достигнуть в Германии в 2009 году. Средняя абсолютная ошибка по шести кандидатам составила всего 1,65% [7].

В дальнейшем к простому подсчёту упоминаний подключился и сентимент анализ. Описание исследований с использованием вышеописанной технологии можно встретить в статье под смелым названием «Кому нужны опросы?» [3] Дэвида Каммингса. Подобным образом в 2010 году статистически значимые данные на уровне 4,34% +/- 2,13% погрешности получил Василеос Лампос [5], предсказав распределение голосов между тремя основными партиями Великобритании.

В попытке обойти очевидный недостаток методов на основе социальных сетей - смещение выборки, было установлено влияние частоты использования ресурса пользователем на прогностическую способность его сообщений [2]. Дифференциация пользователей на «молчаливое большинство» и «активное меньшинство» показала, что мнение последних следует ценить больше, их способности в прогнозировании результатов выборов выше, чем у «молчунов» [2]. Подобное сегментирование, по мнению авторов, помогает произвести более точный отбор сообщений от истинных избирателей и присвоить им соответствующую значимости, что увеличивает точность полученных результатов.

Однако, сомнение вызывает прозрачность, простота и повторяемость методов сбора и анализа данных в социальных сетях. На помощь скептикам приходят данные исследований, эмпирически подтверждающие несовершенства метода. Исследование Гайо-Авелло [4] показало, что настроение твиттер сообщений о работе президента США Барака Обамы не имело положительной корреляции с электоральными предпочтениями. Неудачу ученый объяснил несовершенством упрощенного машинного сентимент анализа. Одно из исследований 2008 года показало, что сентимент анализ серии данных (так называемые time-series) слишком изменчив, для того чтобы иметь возможно предсказать результаты выборов [1].

Также со скептицизмом Гайо-Авелло относится к способам корректировки смещения выборки (один из которых был представлен выше), он полагает, что: «...до тех пор пока социальные медиа не будут регулярно использоваться подавляющим большинством людей, его пользователи не могут считаться репрезентативной выборкой и прогнозы от таких данных, в лучшем случае, будут иметь сомнительную ценность» [Гайо-Авелло, 2011, с.14 (перевод автора)].

В отношении любого исследования социальных сетей встает вопрос, о чем же свидетельствуют полученные данные, кого они способны репрезентировать. Часть населения исключены из производства онлайн данных из-за бедности, не говоря уже о тех, кто имея возможности, не хочет выставлять свои мысли и идеи на общественное обозрение. Появление дешевых технологий может уменьшить цифровой разрыв, расширив круг пользователей, однако, говорить о полной замене традиционных методов в политических прогнозах пока рано. Методики анализа социальных сетей требуют совершенствования, а также эмпирического подтверждения повторяемости, на данный момент большинство успешных прогнозов могут оказаться не более чем случайностью.

### Источники и литература

- 1) Bravo-Marquez, F. Opinion Dynamics of Elections in Twitter/ F. Bravo-Marquez, D. Gayo-Avello, M. Mendoza, B. Pollute // IEEE Computer Society Washington, DC, USA, 2012. P. 32–39.
- 2) Chen, L. Are Twitter Users Equal in Predicting Elections? A Study of User Groups in Predicting 2012 US Republican Presidential Primaries / L. Chen, W. Wang, A. P. Sheth // Social Informatics. Springer. 2012. P. 379–392.
- 3) Cummings, D. Who Needs Polls? Gauging Public Opinion from Twitter Data / D. Cummings, H. Oh, N. Wang// Social Science Computer Review Volume 33 Issue 6, Thousand Oaks, CA, USA. 2015. P. 712-729.
- 4) Gayo-Avello, D. Don't Turn Social Media into Another 'Literary Digest' Poll // Communications of the ACM №54.10. 2011. P.121.
- 5) Lampos, Vasileios On Voting Intentions Inference from Twitter Content: a Case Study on UK 2010 General Election // arXiv preprint arXiv:1204.0423, New York, USA. 2012.
- 6) Smith, A. Twitter Use 2012 / A. Smith, J. Brenner // Pew Internet & American Life Project, Washington, DC, USA. 2012. P. 2-12.
- 7) Tumasjan, A. Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape / A. Tumasjan, T. O. Sprenger, Timm, P.G. Sandner, I. M. Welpе // Social Science Computer Review, Vol. 29. 2010. P. 402–418.
- 8) Trumper, D.S. From Total Hits to Unique Visitors Model for Election's Forecasting / D.S. Trumper, W. Meira, V. Almeida. 2011.

[http://journal.webscience.org/473/1/195\\_paper.pdf](http://journal.webscience.org/473/1/195_paper.pdf)