

Министерство образования и науки Российской Федерации
ФГАОУ ВО «УрФУ имени первого Президента России Б.Н. Ельцина»
Радиоэлектроники и информационных технологий – РТФ
Школа профессионального и академического образования

Оценка работы оценка (зачет/не зачет)

Руководитель от УрФУ _____

ВВЕДЕНИЕ ВКР
по учебной практике

Студент Петросян Александр Артурович

подпись

Специальность Инженерия машинного обучения
09.04.01 Информатика и вычислительная техника
Группа РИМ-231902

Екатеринбург 2025

ВВЕДЕНИЕ

Геолокационные данные в социальных сетях представляют собой ценный источник информации для анализа социальных взаимодействий и пространственно-временных паттернов поведения людей. В последние годы анализ геопривязанных данных из социальных сетей стал важным инструментом для понимания городской динамики, туристических потоков, социальных событий и других аспектов человеческой активности [1,2].

Социальная сеть ВКонтакте, являющаяся одной из крупнейших в России и странах СНГ, содержит огромное количество фотографий с геопривязкой, которые отражают реальные социальные взаимодействия и перемещения пользователей. Эти данные могут быть использованы для выявления закономерностей в пространственно-временной активности людей, популярных мест, сезонных изменений в посещаемости различных локаций и других важных социальных феноменов [3].

Особую ценность представляет возможность анализа данных, собранных на протяжении длительного периода времени. Сервис сбора фотографий, разработанный в 2019-2020 гг., к настоящему моменту накопил базу данных объемом около 150 миллионов объектов, что позволяет проводить масштабные исследования с высокой статистической значимостью. Такие большие объемы данных требуют применения современных методов обработки и анализа, включая технологии Big Data и алгоритмы машинного обучения [4,5].

В отличие от традиционных методов исследования социальной активности, таких как опросы и наблюдения, анализ данных из социальных сетей позволяет получить более объективную и репрезентативную картину, не подверженную субъективным искажениям. Исследования Huang et al. [6] и Lansley et al. [7] показывают, что геопривязанные данные из социальных сетей могут служить надежным индикатором реальных социальных процессов при

правильном методологическом подходе и учете потенциальных смещений выборки.

Как отмечают Li et al. [8], геолокационные данные из социальных сетей могут быть использованы для решения таких практических задач, как:

- оптимизация городской инфраструктуры;
- планирование транспортных маршрутов;
- выявление туристических аттракторов;
- анализ событийной активности;
- оценка влияния природных и социальных явлений на перемещения людей;
- маркетинговые исследования и территориальное планирование бизнеса.

Данная работа направлена на разработку методологии анализа пространственно-временной динамики социальных взаимодействий на основе геопривязанных фотографий из социальной сети ВКонтакте и выявление закономерностей в социальной активности пользователей.

Цель ВКР – разработать методику анализа пространственно-временной динамики социальных взаимодействий на основе геопривязанных фотографий из социальной сети ВКонтакте и оценить применимость различных методов машинного обучения для выявления закономерностей и паттернов в собранных данных.

Для достижения этой цели были поставлены следующие задачи:

1. Провести анализ существующих методов и подходов к исследованию геопространственных данных из социальных сетей.
2. Разработать методологию предварительной обработки и фильтрации данных для повышения их качества и репрезентативности.

3. Определить оптимальные методы пространственного анализа для выявления кластеров активности и популярных локаций.

4. Разработать подход к временному анализу данных для выявления паттернов и трендов в активности пользователей.

5. Исследовать возможности применения методов социального анализа для выявления связей между пользователями и характера их взаимодействия.

6. Интегрировать результаты различных типов анализа для получения комплексного представления о пространственно-временной динамике социальных взаимодействий.

Результатом работы будет методология анализа геопространственных данных из социальных сетей, которая позволит исследователям и практикам в области социологии, урбанистики и маркетинга получать ценные инсайты о социальной активности и пространственном поведении людей.

1 Анализ поставленной задачи и выбор методологии

Задача анализа пространственно-временной динамики социальных взаимодействий на основе геопривязанных фотографий может быть рассмотрена с нескольких ключевых перспектив. Во-первых, это пространственный анализ, направленный на выявление кластеров активности, популярных локаций и пространственных паттернов. Во-вторых, это временной анализ, позволяющий отследить сезонные изменения, суточные ритмы и динамику популярности различных мест со временем. В-третьих, это социальный анализ, направленный на выявление связей между пользователями, социальных групп и характера взаимодействия в различных локациях.

1.1 Обзор существующих подходов к анализу геопространственных данных

Существует несколько ключевых подходов к анализу геопространственных данных из социальных сетей, которые были исследованы

в работах различных авторов. Steiger et al. [9] предлагают комплексный фреймворк для анализа геопривязанных данных из Twitter, который включает в себя этапы сбора, фильтрации, агрегации, анализа и визуализации данных. Авторы отмечают важность учета пространственных, временных и тематических аспектов при анализе таких данных.

В работе Hasan et al. [10] представлен подход к анализу пространственно-временных паттернов городской мобильности на основе данных из Foursquare и Twitter. Авторы используют методы кластеризации и классификации для выявления типичных траекторий перемещения пользователей и факторов, влияющих на выбор маршрутов.

Исследование Chen et al. [11] фокусируется на применении методов глубокого обучения для анализа геопривязанных фотографий из Flickr и Instagram. Авторы предлагают архитектуру нейронной сети, которая позволяет одновременно учитывать пространственные, временные и визуальные характеристики фотографий для предсказания популярности различных локаций.

1.2 Выбор методов для проведения исследования

Для проведения данного исследования необходимо использовать комбинацию методов из области геоинформатики, анализа временных рядов и машинного обучения. Пространственный анализ может быть проведен с использованием методов кластеризации, таких как DBSCAN или HDBSCAN, которые хорошо подходят для выявления кластеров произвольной формы в геопространственных данных [12,13]. Временной анализ может быть проведен с использованием методов декомпозиции временных рядов, спектрального анализа и прогнозирования [14]. Социальный анализ может включать в себя построение графов взаимодействия и применение методов анализа социальных сетей [15].

Сравнительный анализ различных методов кластеризации для геопространственных данных, проведенный Virant и Kut [16], показывает, что DBSCAN и его модификации обеспечивают наилучшие результаты для данных с неравномерным распределением плотности, что характерно для геопривязанных фотографий из социальных сетей. Метод HDBSCAN, предложенный Campello et al. [17], является усовершенствованным вариантом DBSCAN, который способен автоматически определять оптимальное количество кластеров и работать с данными различной плотности.

Для анализа временных рядов могут быть использованы такие методы, как сезонная декомпозиция STL (Seasonal-Trend decomposition using LOESS), предложенная Cleveland et al. [18], или более современные подходы, такие как Prophet, разработанный Facebook [19], которые позволяют выявлять тренды, сезонность и аномалии во временных данных.

В отличие от традиционных исследований геолокационных данных, которые часто фокусируются только на пространственном аспекте, данная работа предполагает интегрированный подход, учитывающий все три измерения: пространство, время и социальные взаимодействия. Это позволит получить более полную и многогранную картину социальной активности, как это было продемонстрировано в работах Hawelka et al. [20] и Silva et al. [21].

1.3 Структура и особенности исходных данных

Собранная база данных представляет собой структурированный набор информации о фотографиях с геопривязкой, включающий такие атрибуты, как:

- имя файла: уникальный идентификатор фотографии
- координаты (долгота и широта): геопространственное положение фотографии с точностью до нескольких метров
- дата и время: временная метка создания фотографии

- ссылка на изображение: URL для доступа к полноразмерному изображению
- ссылка на миниатюру: URL для доступа к миниатюре изображения
- оценка: числовой показатель популярности фотографии
- дополнительные метаданные: теги, описания, информация о пользователе (с учетом ограничений приватности)

Эта структура данных позволяет проводить разнообразные типы анализа, от простого картографирования до сложных методов машинного обучения. По своим характеристикам данные соответствуют понятию "больших данных" (Big Data), так как обладают значительным объемом (Volume), разнообразием (Variety) и высокой скоростью обновления (Velocity) [22].

1.4 Методы машинного обучения для анализа данных

Ключевой подход в данной работе будет основан на методах машинного обучения, которые позволяют выявлять неочевидные закономерности в больших объемах данных. Для анализа пространственных данных будут использованы методы кластеризации и классификации, для временного анализа – модели прогнозирования и выявления аномалий, а для социального анализа – методы анализа графов и сетей.

Yang et al. [23] демонстрируют эффективность применения алгоритмов глубокого обучения, таких как сверточные нейронные сети (CNN) и рекуррентные нейронные сети (RNN), для анализа геопространственных данных. Авторы показывают, что такие модели способны учитывать как пространственные, так и временные зависимости в данных и обеспечивают более высокую точность предсказаний по сравнению с традиционными методами.

Для выявления аномалий и необычных паттернов в данных могут быть использованы такие методы, как изолирующий лес (Isolation Forest) [24] или

одноклассовый SVM (One-Class SVM) [25], которые способны идентифицировать наблюдения, значительно отличающиеся от общей тенденции.

2 Методология анализа геопространственных данных из социальных сетей

Методология анализа геопространственных данных из социальных сетей включает в себя несколько ключевых этапов: сбор и предварительная обработка данных, пространственный анализ, временной анализ, социальный анализ и интерпретация результатов.

2.1 Сбор и предварительная обработка данных

На этапе предварительной обработки данных производится фильтрация ошибочных или неполных записей, нормализация координат, преобразование временных меток в удобный для анализа формат и обогащение данных дополнительной информацией, такой как тип местности, административная принадлежность и т.д.

Важным аспектом предварительной обработки является также фильтрация дубликатов и ботов, которые могут исказить результаты анализа. Методы обнаружения ботов в социальных сетях, предложенные Minnich et al. [26] и Ferrara et al. [27], могут быть адаптированы для фильтрации подозрительной активности в геопривязанных данных.

Для повышения качества и информативности данных может быть проведено их обогащение с использованием внешних источников информации, таких как OpenStreetMap [28] или Google Places API [29], которые позволят добавить к геопривязанным фотографиям информацию о типе объектов, расположенных в их окрестности (рестораны, парки, музеи и т.д.), что может быть полезно для последующего анализа.

2.2 Пространственный анализ

Пространственный анализ включает в себя визуализацию плотности точек на карте с использованием тепловых карт (heatmaps), выявление кластеров с помощью алгоритмов кластеризации (K-means, DBSCAN, HDBSCAN), анализ пространственной автокорреляции с использованием индекса Морана и выявление пространственных паттернов с помощью методов машинного обучения.

Тепловые карты могут быть построены с использованием методов оценки плотности ядра (Kernel Density Estimation, KDE) [30], которые позволяют визуализировать распределение интенсивности активности в пространстве. Для анализа пространственной автокорреляции может быть использован глобальный и локальный индекс Морана [31], который позволяет выявить области с высокой или низкой концентрацией активности и их пространственные взаимосвязи.

Для выявления значимых мест и регионов может быть использован метод пространственно-временного сканирования (Space-Time Scan Statistics) [32], который позволяет идентифицировать кластеры активности, статистически значимо отличающиеся от фонового уровня.

2.3 Временной анализ

Временной анализ включает в себя декомпозицию временных рядов для выявления тренда, сезонности и остаточной компоненты, анализ суточных, недельных и сезонных паттернов активности, выявление аномалий и изменений в активности со временем.

Для анализа временных паттернов может быть использован метод спектрального анализа [33], который позволяет выявить основные периодические компоненты во временных рядах. Для выявления трендов и сезонности может быть применен метод STL (Seasonal-Trend decomposition using

LOESS) [18] или более современный метод Prophet [19], разработанный специально для анализа временных рядов с ярко выраженной сезонностью.

Для выявления аномалий и необычных событий во временных рядах могут быть использованы методы, основанные на статистическом подходе, такие как метод GESD (Generalized Extreme Studentized Deviate) [34], или методы машинного обучения, такие как автоэнкодеры [35] или изолирующий лес [24].

2.4 Социальный анализ

Социальный анализ направлен на выявление социальных групп и связей между пользователями, анализ характера взаимодействия в различных локациях и временных промежутках, а также изучение влияния социальных факторов на пространственно-временные паттерны.

Для анализа социальных связей может быть использован метод построения и анализа графов взаимодействия [36], где узлами являются пользователи, а ребрами – связи между ними, определяемые на основе совместного присутствия в одних и тех же локациях или других признаков социального взаимодействия.

Для выявления социальных групп может быть применен метод обнаружения сообществ в графах [37], такой как алгоритм Лувена (Louvain) или алгоритм InfoMap, которые позволяют идентифицировать группы пользователей с высокой степенью взаимосвязи.

Для анализа влияния социальных факторов на пространственно-временные паттерны может быть использован метод регрессионного анализа [38] или методы машинного обучения, такие как случайный лес (Random Forest) [39] или градиентный бустинг (Gradient Boosting) [40].

2.5 Интеграция результатов и визуализация

Для интеграции результатов различных типов анализа предлагается использовать методы многомерного анализа, такие как факторный анализ, многомерное шкалирование и методы уменьшения размерности (PCA, t-SNE), которые позволяют выявить скрытые закономерности в многомерных данных.

Метод главных компонент (Principal Component Analysis, PCA) [41] позволяет сократить размерность пространства признаков, сохраняя при этом максимум информации о вариативности данных. Метод t-SNE (t-distributed Stochastic Neighbor Embedding) [42] может быть использован для визуализации многомерных данных в двумерном или трехмерном пространстве, сохраняя структуру локального соседства.

Для визуализации результатов анализа могут быть использованы интерактивные карты и графики, созданные с помощью библиотек D3.js [43], Plotly [44] или Tableau [45]. Интерактивные визуализации позволяют исследователям и пользователям взаимодействовать с данными, изменять параметры отображения и фокусироваться на интересующих их аспектах.

2.6 Этические аспекты и вопросы приватности

Важным аспектом методологии является также учет этических и privacy-ориентированных аспектов работы с данными из социальных сетей. Хотя все данные являются публично доступными, необходимо обеспечить анонимизацию и агрегацию данных на уровне, который исключает возможность идентификации конкретных пользователей.

Как отмечают Zimmer [46] и Zook et al. [47], работа с геопривязанными данными из социальных сетей требует особого внимания к вопросам приватности, так как такие данные могут потенциально раскрывать чувствительную информацию о пользователях, такую как места их проживания, работы и отдыха.

Для обеспечения приватности в данной работе будут использованы методы k-анонимизации [48] и дифференциальной приватности [49], которые позволяют защитить личную информацию пользователей при сохранении статистической значимости результатов анализа.

2.7 Оценка результатов и валидация

Для оценки результатов анализа предлагается использовать методы валидации моделей машинного обучения, такие как кросс-валидация, а также методы оценки качества кластеризации, такие как силуэтный коэффициент и индекс Дэвиса-Болдина.

Метод кросс-валидации [50] позволяет оценить обобщающую способность моделей машинного обучения путем разделения данных на обучающую и тестовую выборки в различных конфигурациях. Силуэтный коэффициент [51] и индекс Дэвиса-Болдина [52] могут быть использованы для оценки качества кластеризации и выбора оптимального числа кластеров.

Для валидации результатов временного анализа могут быть использованы метрики, такие как MAE (Mean Absolute Error), RMSE (Root Mean Square Error) и MAPE (Mean Absolute Percentage Error) [53], которые позволяют оценить точность прогнозов и адекватность моделей временных рядов.

Результаты анализа будут представлены в виде интерактивных визуализаций, карт и графиков, которые позволят наглядно продемонстрировать выявленные закономерности и паттерны в пространственно-временной динамике социальных взаимодействий.

Список литературы:

1. Huang, Q., Wong, D. W. (2015). Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data. *Annals of the Association of American Geographers*, 105(6), 1179-1197.
2. García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, 63, 408-417.
3. Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal*, 78(2), 319-338.
4. Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171-209.
5. Zhou, X., Xu, C., & Kimmons, B. (2015). Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Information Systems Frontiers*, 17(6), 1239-1252.
6. Huang, Q., Cao, G., & Wang, C. (2014). From where do tweets originate?: A GIS approach for user location inference. In *Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks* (pp. 1-8).
7. Lansley, G., & Longley, P. A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems*, 58, 85-96.
8. Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, 40(2), 61-77.
9. Steiger, E., de Albuquerque, J. P., & Zipf, A. (2015). An advanced systematic literature review on spatiotemporal analyses of Twitter data. *Transactions in GIS*, 19(6), 809-834.
10. Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing* (pp. 6:1-6:8).

11. Chen, M., Liu, Y., & Yu, X. (2015). NLPMM: A next location predictor with Markov modeling. In Proceedings of the 19th Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 186-197).
12. Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (pp. 226-231).
13. Schubert, E., Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. ACM Transactions on Database Systems, 42(3), 19:1-19:21.
14. López-de-Lacalle, J. (2019). tsoutliers: Detection of outliers in time series. R package version. <https://CRAN.R-project.org/package=tsoutliers>
15. Barabási, A.-L. (2016). Network Science. Cambridge University Press.
16. Birant, D., & Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. Data & Knowledge Engineering, 60(1), 208-221.
17. Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 160-172).
18. Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics, 6(1), 3-73.
19. Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37-45.
20. Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. Cartography and Geographic Information Science, 41(3), 260-271.
21. Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Salles, J., & Loureiro, A. A. F. (2013). A picture of Instagram is worth more than a thousand words: Workload characterization and application. In Proceedings of the IEEE International Conference on Distributed Computing in Sensor Systems (pp. 123-132).

22. Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. META Group Research Note, 6(70), 1.
23. Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., & Moscibroda, T. (2016). Mobility modeling and prediction in bike-sharing systems. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (pp. 165-178).
24. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In Proceedings of the 8th IEEE International Conference on Data Mining (pp. 413-422).
25. Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. Neural Computation, 13(7), 1443-1471.
26. Minnich, A., Chavoshi, N., Koutra, D., & Mueen, A. (2017). BotWalk: Efficient adaptive exploration of Twitter bot networks. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 467-474).
27. Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104.
28. Haklay, M., & Weber, P. (2008). OpenStreetMap: User-generated street maps. IEEE Pervasive Computing, 7(4), 12-18.
29. Google Places API. (2021). Retrieved from <https://developers.google.com/maps/documentation/places/web-service/overview>
30. Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall/CRC.
31. Anselin, L. (1995). Local indicators of spatial association - LISA. Geographical Analysis, 27(2), 93-115.
32. Kulldorff, M. (1997). A spatial scan statistic. Communications in Statistics - Theory and Methods, 26(6), 1481-1496.
33. Stoffer, D. S. (1999). Detecting common signals in multiple time series using the spectral envelope. Journal of the American Statistical Association, 94(448), 1341-1356.

34. Rosner, B. (1983). Percentage points for a generalized ESD many-outlier procedure. *Technometrics*, 25(2), 165-172.
35. Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). Long short term memory networks for anomaly detection in time series. In *Proceedings of the 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (pp. 89-94).
36. Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.
37. Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5), 75-174.
38. Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis* (3rd ed.). Wiley-Interscience.
39. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
40. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232.
41. Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer.
42. van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579-2605.
43. Bostock, M., Ogievetsky, V., & Heer, J. (2011). D³: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2301-2309.
44. Plotly Technologies Inc. (2015). Collaborative data science. Plotly Technologies Inc. <https://plot.ly>
45. Tableau Software. (2021). Tableau. <https://www.tableau.com/>
46. Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology*, 12(4), 313-325.
47. Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., ... & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLOS Computational Biology*, 13(3), e1005399.

48. Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5), 557-570.
49. Dwork, C. (2006). Differential privacy. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming* (pp. 1-12).
50. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 1137-1143).
51. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.
52. Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224-227.
53. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679-688.