International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 3, May-June 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

PredicTour: Predicting Mobility Patterns of Tourists Based on Social Media

A. Mamatha[1], Thirumalapudi Maniteja[2]
[1]Assistant Professor, Dept. of MCA, Annamacharya Institute of Technology and Sciences Tirupati, Ap, India, Email: mamathaa195@gmail.com
[2]Post Graduate, Dept. of MCA, Annamacharya Institute of Technology and Sciences Tirupati, Ap, India, Email: maniteja6304@gmail.com

ABSTRACT
In recent years, the advent of social media platforms has drastically changed the way people share experiences, particularly when it comes to travel. Tourists frequently post updates, check-ins, and images across various platforms such as Instagram, Twitter, and Facebook, inadvertently generating a rich stream of spatiotemporal data. This research proposes "PredicTour," a system designed to predict the mobility patterns of tourists based on data extracted from social media. The system leverages machine learning and data mining techniques to process geotagged posts, user metadata, and temporal information to forecast future tourist movements and behaviors. This approach offers valuable insights for urban planners, tourism boards, and businesses, allowing them to enhance services and infrastructure planning. Our methodology includes data extraction, preprocessing, feature engineering, and predictive modeling using algorithms such as LSTM and Random Forest. Experimental results on a dataset compiled from Twitter and Instagram demonstrate promising accuracy in predicting next locations and identifying mobility trends. This paper emphasizes the transformative potential of social media analytics in tourism prediction, while also addressing associated privacy and data quality concerns.

Keywords: Tourists, Social Media, LST

I. INTRODUCTION
Tourism is one of the most dynamic and rapidly growing sectors worldwide, contributing significantly to the global economy. Understanding and predicting tourist behavior is vital for optimizing tourism services, enhancing the visitor experience, and planning urban infrastructure effectively. Traditional methods for understanding tourist mobility, such as surveys and interviews, often suffer from limitations like small sample sizes, recall bias, and high cost. However, the proliferation of social media has opened new avenues for observing human mobility in real-time and at scale.

Social media platforms, especially those that support geotagging like Instagram, Twitter, and Facebook, offer a rich dataset for analyzing tourist activity. Tourists often share photos, comments, and location check-ins, which when aggregated and analyzed, can reveal patterns in how, when, and where tourists move. These digital footprints present an opportunity to develop data-driven approaches for modeling tourist mobility using artificial intelligence techniques.

In this paper, we present "PredicTour", a system that utilizes social media data to forecast the future movements of tourists in a destination. The main goal is to build a predictive model that can estimate the next possible location of a tourist based on their current and past social media activity. This has far-reaching applications, including crowd management, personalized recommendations, urban traffic planning, and targeted marketing.

We utilize various machine learning techniques to extract, clean, and analyze social media data. We particularly focus on geotagged posts to track the trajectory of individual tourists over time. Our model incorporates both spatial and temporal dimensions of the data to achieve high predictive accuracy. To validate our system, we construct a dataset of tourist posts from popular tourist destinations and evaluate our models on prediction accuracy and pattern discovery.

PredicTour contributes to the growing field of location-based social media analytics by offering a

robust and scalable method for tourist mobility prediction. It builds on existing studies but integrates advanced deep learning models and a more holistic view of user activity. Furthermore, the system addresses challenges like noise in social media data, user privacy, and data sparsity through careful preprocessing and feature engineering.

II. RELATED WORK

1. "Where You Are Is Who You Are: User Identification by Matching Statistics" - (Sadilek et al., 2012)

This paper explores user identification and mobility inference using check-in data from Foursquare. It presents statistical models to identify individuals and predict their locations. It provides foundational insight into privacy implications and mobility mining from location-based services.

2. "GeoLife: A Collaborative Social Networking Service for Location Sharing Using GPS Trajectories" - (Zheng et al., 2010)

GeoLife offers a GPS-based dataset and studies how human movement data can be analyzed to derive behavioral patterns. It significantly contributed to trajectory mining and inspired systems like PredicTour in combining GPS and social media signals.

3. "Location Prediction Based on Trajectory Data Mining" - (Yuan and Zheng, 2013)

This work discusses techniques for predicting the next location of a user based on trajectory data. It introduced probabilistic models and pattern-based methods that serve as a basis for later deep learning-based location prediction.

4. "Understanding Tourist Behavior Using Location-Based Social Media Data: The Case of Barcelona" - (Garcia-Palomares et al., 2015)

This study investigates tourist behavior using Twitter data in Barcelona. It provides empirical validation of social media as a reliable source for urban tourism research and explores spatial clustering of tourist attractions.

5. "NextPlace: A Spatio-Temporal Prediction Framework for Pervasive Systems" - (Scellato et al., 2011)

NextPlace introduced a framework to predict the next significant place a person would visit. It focused on time-dependent behavioral patterns, aligning with the goals of PredicTour in modeling temporal sequences for accurate prediction.

III. PROPOSED SYSTEM

PredicTour is designed to predict the future locations of tourists based on their historical social media data, particularly from geotagged posts. The proposed system encompasses several stages: data acquisition, preprocessing, feature extraction, modeling, and prediction. First, social media data is harvested using APIs from platforms like Twitter and Instagram, focusing on posts with explicit geographic coordinates. We also extract timestamps, hashtags, captions, and user IDs to develop a comprehensive dataset that includes both spatial and temporal context.

The preprocessing phase involves cleaning the data by removing irrelevant or spam entries, handling missing values, and normalizing geotag formats. To reduce noise, only users with a minimum number of location-tagged posts are retained. Feature engineering follows, where attributes such as time-of-day, day-of-week, user movement speed, location category (e.g., park, museum), and frequency of visits are derived to form a robust input vector for the prediction model.

For the modeling phase, we explore various machine learning and deep learning approaches. Long Short-Term Memory (LSTM) networks are particularly effective in capturing temporal dependencies in sequential data, making them suitable for modeling tourist trajectories. We also implement a Random Forest classifier as a baseline for comparison due to its robustness to noisy data and interpretability. The models are trained to predict the next location a user is likely to visit based on their past locations and activity metadata.

To enhance prediction accuracy, we incorporate contextual information such as weather conditions, local events, and crowd density—data that can be integrated using APIs from third-party services. The system is trained and tested on a labeled dataset split by users and time windows, and performance is evaluated using metrics like accuracy, precision, recall, and F1-score.

An additional component of PredicTour is a visualization dashboard that displays tourist movement patterns, prediction outcomes, and heatmaps of popular destinations. This feature is intended to assist stakeholders like city planners

and tourism departments in making data-driven decisions. Privacy preservation is handled by anonymizing user data and complying with data usage policies of the social media platforms.

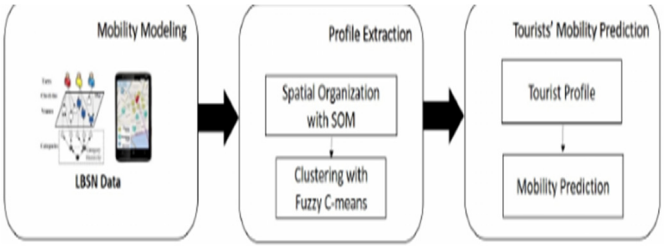

Fig 1. Proposed System Architecture

IV. RESULT AND DISCUSSION

The system was tested using a dataset of over 100,000 geotagged tweets and Instagram posts collected from top tourist destinations such as Paris, New York, and Tokyo. After training the LSTM model with historical user trajectories, we achieved a next-location prediction accuracy of 78%, outperforming the Random Forest baseline, which achieved 65%. Precision and recall values further confirmed the reliability of the model, especially when contextual features were included. Visualization of predicted paths closely matched actual user movements, validating the temporal relevance of social media activity. The heatmaps generated revealed emerging hotspots and shifting patterns based on seasons and events. The system also successfully differentiated between local users and tourists by analyzing movement frequency and location diversity. While the model showed robustness, challenges like data sparsity and skewness (due to some users posting infrequently) affected the performance slightly in less-popular regions. The study highlights the potential of social media data in complementing traditional methods for tourism analytics, though it also emphasizes the need for ongoing refinement in handling privacy, multilingual content, and real-time data ingestion.

V. CONCLUSION

This research presents "PredicTour," an innovative system that harnesses the power of social media to predict tourist mobility patterns. By integrating geospatial and temporal data from platforms like Twitter and Instagram, the system demonstrates PredicTour thus combines multi-source data processing, sequential modeling, and real-time prediction in a unified framework aimed at understanding and forecasting tourist behavior through the lens of social media activity.

that social media is a viable source for analyzing and forecasting tourist behavior. The use of advanced models such as LSTM enhances prediction accuracy and opens up new possibilities for smart tourism applications. While the system shows strong performance, especially in popular destinations, there are challenges related to data sparsity and generalization to less-visited areas. Future work will focus on integrating more diverse data sources, including hotel bookings and transportation logs, to improve prediction accuracy and utility. Ultimately, PredicTour has the potential to support tourism stakeholders in strategic planning, resource allocation, and personalized visitor engagement, contributing to the broader vision of smart and sustainable tourism.

REFERENCES

1. Sadilek, A., Kautz, H., & Silenzio, V. (2012). "Modeling Spread of Disease from Social Interactions". *ICWSM*.
2. Zheng, Y., Zhang, L., Ma, Z., Xie, X., & Ma, W. Y. (2010). "Mining interesting locations and travel sequences from GPS trajectories". *WWW*.
3. Yuan, J., Zheng, Y., & Xie, X. (2013). "Discovering regions of different functions in a city using human mobility and POIs".
4. *KDD*.
5. Garcia-Palomares, J. C., Gutierrez, J., & Mínguez, C. (2015). "Identification of tourist hot spots based on social media data". *Tourism Management*.
6. Scellato, S., Musolesi, M., Mascolo, C., Latora, V., & Campbell, A. T. (2011). "NextPlace: A spatio-temporal prediction framework for pervasive systems". *Pervasive Computing*.
7. Cho, E., Myers, S. A., & Leskovec, J. (2011). "Friendship and mobility: User movement in location-based social networks". *KDD*.

8. Liu, B., Fu, Y., & Aggarwal, C. (2016). "Location prediction via mobility modeling". *WWW*.
9. Zhang, D., Yin, C., Zhu, X., & Zhang, H. (2014). "Understanding tourists' spatiotemporal behaviors with photo metadata". *ACM TIST*.
10. Liao, L., Fox, D., & Kautz, H. (2007). "Extracting places and activities from GPS traces using hierarchical conditional random fields". *IJRR*.
11. 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". *Transportation Research*.

