Journal of Engineering Sciences
Vol 11, Issue 5,May/2020
ISSN NO:0377-9254

ANALYTICS OF TOURISM ATTRACTION ROUTES USING TRANSPORT DATA

G. Bhargava Deep [1], N. Sruthi [2], V. Surya Vamsi [3], Dr. Riyazuddin Y Md [4]
[1,2,3] *Department of Computer Science And Engineering, GITAM (Deemed to be University), Hyderabad, India.*

[4] *Assistant Professor, Department of Computer Science And Engineering, , GITAM (Deemed to be University), Hyderabad, India.*

***Abstract*— In this project we would address the issue of identifying tourists out of large group of public at a particular tourist place. We will develop a framework which lists out a complete database which contains various details about the place, its attractions, its approximate costs and its limitations. At present the existing system uses the social networking data and transport data to make predictions. The disadvantage of this system that it has limited coverage and unpredictable information delay. To overcome these limitations, we propose a new system which gives a detailed information of the tourist place and the various details of the place such as its connectivity from other places, cost to reach, modes of transport, review and feedback. Therefore the proposed system helps a tourist to find all the information related to a place he/she is interested to visit and makes it easier to visit and recommend it to another tourist by posting his reviews and feedback about the place.**

***Keywords*— Tourism, Transport data, analytics, routes**

1. INTRODUCTION

In this project, we addressed the issue of segregating the foreign tourists from the database containing a whole list of tourists. We also solve the problems occured due to relying on social data to make predictions of the next tours of the tourists. This project recommends, to the tourists, their next tours based on their previous tourism history and on the basis of the type of tourist(foreign or domestic) he/she is. We also achieve an efficient and hassle free approach towards finding information on the monuments. We also find, suggest the best and minimal possible route to the specified monument. For this we take the existing system, Tour Sense: A Framework for Tourist Identification and Analytics Using Transport Data and try to solve the problem with this format. To capture and understand tourists and their preferences, the recent tourism analytics research mainly adopts social media data (e.g., geotagged images in Flickr) [1], [2], [3], where the basic assumption behind this attempt isthat most tourists would like to share their travel moments on their online social networks. However, using social media data may suffer from the limited coverage and information delay: (a) only a small portion of tourists are actively sharing their photos or travel experiences on social media, as many travellers may not be the fans of social networks or even not use the Internet. Furthermore, most shared contents are popular landmarks, not covering all the places a tourist visited, and thus the insight gained from social media data may be incomplete or biased; (b) considering the high data roaming fees, many social network sharings are not real-time posted. Tourists may share their photos and feelings after a whole day's travel, or even after coming back to their hometowns. Meanwhile, how to effectively and timely crawl all the tourists' social media information from the service providers is also challenging. Besides the social media data, sensor network data (e.g., bluetooth data) [4] and cellular data [5] are also adopted by the researchers for tourist study, but they suffer from the similar limitations and

constraints.This work attempts to tackle the above issues, by demonstrating how the transport data can be used to identify and analyze tourists. Despite of a diversity of local tour services available, public transport (e.g., metro and bus) is still the most cost-efficient and convenient travelling approach for most tourists, especially in the densely-populated cities like Singapore and Tokyo. Accordingly, the public transport data offer a sufficient coverage of the tourist population. Meanwhile, the widely adopted electronic fare payment systems can timely record and trace tourists and their travelling routes, when they tap in/ out at the gantry of a station or boarding/alighting on a bus. In particular, we propose a novel but practical framework for tourist analytics, called TourSense, that (a) first applies machine learning techniques on transport data to identify tourists from public commuters, and (b) uses the identified tourist travelling information to conduct their preference analytics and thereby timely makes the personalized recommendation and prediction. To provide the practical embodiments of the proposed framework, we take Singapore as an exemplary case and present the empirical experiment results using the public transport data from the city. Identification on Tourists from Public Commuters: Using the transport data, we propose a two-phase algorithm to identify tourists from public commuters. The key innovations include (i) properly ranking transport stations according to how they are likely to be a destination for tourists; and (ii) designing a graph-based novel iterative learning algorithm to accomplish the tourist identification. Tourist Preference Analytics: Using the identified tourists and their travel records, we design the personalized preference analytics and location recommendation methods for tourists. The key innovation include (i) a tourist-location transition frequency matrix and a location-location transition frequency matrix are designed to represent the tourist information, and (ii) a novel recommendation model is designed to learn tourists' preferences for individual locations and tours. To the best of our knowledge, this is the first work that analyzes tourists' public transport trajectories for location preference study.Since the Singapore transport data is not available, we decided to use the data set : Monuments in India and list of tourists. In this list of tourists there are both foreigners and the domestic people. So first we segregate them into Foreign Tourists and Domestic Tourists.

2. LITERATURE SURVEY

As one of the world's largest industries, tourism serves as the economic pillar of many countries and cities. The total contribution of the tourism industry to GDP was 7,600 billion U.S. dollars (3.1 percent of global GDP) and supported 292 million jobs (9.6 percent of total employment) in 2016.1 Taking Singapore as an example, its tourism industry brought in more than 16.4 million of foreign tourists (more than thrice the country's population) and created more than 160 thousand jobs for local residents in 2017. Tracking and understanding tourists would directly benefit local government and tour agencies to design and improve their services, such as launching new tour routes and providing customized tour packages based on tourist's characteristics and preferences. To capture and understand tourists and their preferences, the recent tourism analytics research mainly adopts social media data (e.g., geotagged images in Flickr) [1], [2], [3], where the basic assumption behind this attempt isthat most tourists would like to share their travel moments on their online social networks. However, using social media data may suffer from the limited coverage and information delay: (a) only a small portion of tourists are actively sharing their photos or travel experiences on social media, as many travellers may not be the fans of social

networks or even not use the Internet. Furthermore, most shared contents are

popular landmarks, not covering all the places a tourist visited, and thus the insight gained from social media data may be incomplete or biased; (b) considering the high data roaming fees, many social network sharings are not real-time posted. Tourists may share their photos and feelings after a whole day's travel, or even after coming back to their hometowns. Meanwhile, how to effectively and timely crawl all the tourists' social media information from the service providers is also challenging. Besides the social media data, sensor network data (e.g., bluetooth data) [4] and cellular data [5] are also adopted by the researchers for tourist study, but they suffer from the similar limitations and constraints.This work attempts to tackle the above issues, by demonstrating how the transport data can be used to identify and analyze tourists. Despite of a diversity of local tour services available, public transport (e.g., metro and bus) is still the most cost-efficient and convenient travelling approach for most tourists, especially in the densely-populated cities like Singapore and Tokyo. Accordingly, the public transport data offer a sufficient coverage of the tourist population. Meanwhile, the widely adopted electronic fare payment systems can timely record and trace tourists and their travelling routes, when they tap in/ out at the gantry of a station or boarding/ alighting on a bus. In particular, we propose a novel but practical framework for tourist analytics, called TourSense, that (a) first applies machine learning techniques on transport data to identify tourists from public commuters, and (b) uses the identified tourist travelling information to conduct their preference analytics and thereby timely makes the personalized recommendation and prediction. To provide the practical embodiments of the proposed framework, we take Singapore as an exemplary case and present the empirical experiment results

using the public transport data from the city.

3. EXISTING SYSTEM

Existing social media data may suffer from the limited coverage and information delay: only a small portion of tourists are actively sharing their photos or travel experiences on social media, as many travellers may not be the fans of social networks or even not use the Internet. Furthermore, most shared contents are popular landmarks, not covering all the places a tourist visited, and thus the insight gained from social media data may be incomplete or biased considering the high data roaming fees, many social network sharings are not real-time posted. Tourists may share their photos and feelings after a whole day's travel, or even after coming back to their hometowns. Meanwhile, how to effectively and timely crawl all the tourists' social media information from the service providers is also challenging. Besides the social media data, sensor network data (e.g., bluetooth data) [4] and cellular data [5] are also adopted by the researchers for tourist study, but they suffer from the similar limitations and constraints.This work attempts to tackle the above issues, by demonstrating how the transport data can be used to identify and analyze tourists. Despite of a diversity of local tour services available, public transport (e.g., metro and bus) is still the most cost-efficient and convenient travelling approach for most tourists, especially in the densely-populated cities like Singapore and Tokyo. Accordingly, the public transport data offer a sufficient coverage of the tourist population. Meanwhile, the widely adopted electronic fare payment systems can timely record and trace

4. PROPOSED SYSTEM

In particular, we propose a novel but practical framework for tourist analytics,

called TourSense, thatfirst applies machine learning techniques on transport data to identify tourists from public commuters, anduses the identified tourist travelling information to conduct their preference analytics and thereby timely makes the personalized recommendation and prediction. To provide the practical embodiments of the proposed framework, we take Singapore as an exemplary case and present the empirical experiment results using the public transport data from the city. Our work in this paper thus makes the following key contributions: Novel Framework for Tourist Analytics: We propose a novel framework that conducts analytics on tourists using transport data. By leveraging on the citywide bus and subway data, we show how the public transport data can provide hard-to-obtain, tourist-specific insights and quantitative results.Identification on Tourists from Public Commuters: Using the transport data, we propose a two- phase algorithm to identify tourists from public commuters.The key innovations includeproperly ranking transport stations according to how they are likely to be a destination for tourists; anddesigning a graph-based novel iterative learning algorithm to accomplish the tourist identification.Tourist Preference Analytics: Using the identified tourists and their travel records, we design the personalized preference analytics and location recommendation methods for tourists.a tourist-location transition frequency matrix and a location-location transition frequency matrix are designed to represent the tourist information, andnovel recommendation model is designed to learn tourists' preferences for individual locations and tours. To the best of our knowledge, this is the first work that analyzes tourists' public transport trajectories for location preference study.

5. METHODOLOGY

The block diagram of the TourSense framework is illustrated in Fig. 1, which mainly consists of three modules, namely public transportation system, tourist identification system and tourist preference analytics system. Briefly speaking, public transportationsystem provides the transportation data and infrastructure information (e.g., subway/bus data and station information). By leveraging on such data and information, tourist identification system recognizes tourists from public commuters. Using the identified tourists and their travelling traces, tourist preference analytics system further investigates their favorite attractions and tours. All the above tourist information and analytics results will be aggregated, via specifically designed user interface and feedback channel, and eventually provide to different stakeholders, typically including transportation operators, government agencies and tourists themselves. We will elaborate the three systems in this section respectively.

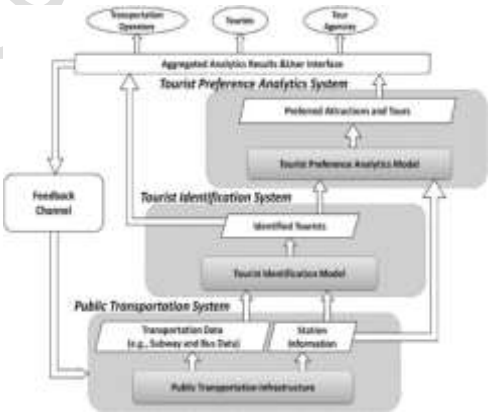

Fig.1 Block diagram of proposed system

5.1. PUBLIC TRANSPORTATION SYSTEM

The public transportation infrastructure covers different urban transportation services (e.g., subway and bus services) and facilities (e.g., subway and bus stations). Each service utilizes their own informatics system to acquire relevant commuter travelling data. For example, today's subway service usually employs RFID-based ticketing system to

automatically collect passengers' ingress and egress (tap-in and tap-out) data at each subway station. The bus service deploys the similar ticketing system on each operating bus to record their boarding and alighting information. The commuter travelling data include both real time transaction data, which can be timely collected by the backend servers of transportation operators, and the historical transaction data, which are usually streamed and stored into a Hadoop distributed file system for daily maintenance and batch processing. In addition, the public transportation system will also provide the station information, including the geographic location and nearby POIs of each metro station or bus stop.

5.2. TOURIST IDENTIFICATION SYSTEM

This system periodically recognizes tourists from commuters using the data and information collected from the public transportation system. More specifically, it targets to identify the transport records that are generated by the riding of tourists from the public transport data. In general, the travelling population can be assumed as two groups, i.e., tourists and non-tourists (non-tourists normally mean local people). Tourists refer to the group of people who visit the city for sightseeing purpose during a short term (e.g., a couple of days). They commonly visit places of interest, including historic sites, museums, restaurants, shopping streets, and stay in hotels or hostels. People who come to the city for other purposes such as business or medical services may not fall into the class of tourists in this system. Some local domain knowledge and a small set of labeled commuters information may be needed during the identification process.

The key outputs of the system is the identified tourist sets and their riding records, which serve as the main inputs of the upper tourist preference analytics system.

5.3. TOURIST PREFERENCE ANALYTICS SYSTEM

Taking advantages of the identified tourist information, especially their travelling traces, this system mainly conducts the preference analytics on the tourists, such as predicting individual tourist's next visiting locations and accordingly making next POI (place of interest) recommendations to those who are not sure about where to go. Such preference analytics results can be utilized in many services. For example, the inferred tourist preferences on his or her unvisited locations can be used to generate the personalized advertisement (e.g., attraction tickets and nearby dining promotions), which can be pushed to the tourists through different feedback channels, such as the screens on the subway station gantry or the top-up machines at the ticketing office. Moreover, the analytics results can be used by the designed user interface to answer "next- visiting-place" queries from tourists.In short, the above-described three systems work cooperatively to acquire, process and analyze the public transportation data for tourists. The final analytics results would possibly benefit different stakeholders, including tourists, transportation operators and tour agencies. We will elaborate our design on tourist identification system and preference analytics systems in the subsequent two sections.

5.4 TOURIST IDENTIFICATION SYSTEM DESIGN

We design a two-phase algorithm to tackle the tourist identification problem. The first phase conducts the so-called station ranking. Its main task is to assign

an initial score to each transportation station that indicating whether it is more likely a destination for tourists or a destination for nontourists. The second phase conducts the so-called iterative propagation learning, where an iterative learning algorithm is designed using the

station ranking results to accomplish the tourist identification task. We will present the two phases in the following parts respectively.

Phase I: Station Ranking

Intuitively, knowing someone who has visited a station with a high (or low) initial score may increase (or reduce) our belief that the person is a tourist. We thus compute a score for each given station to describe whether the station is more likely to be a destination for tourists. However, it is not a proper way to simply use the attractiveness of a place to tourists as the initial scores (such as the scores on the travel sites like TripAdvisor). It is mainly because one place that is popular to tourists may also be popular to locals. For example, most tourists may visit famous shopping streets in a city, while local people may favorite them as well. We thus need to consider the popularity of a place to both tourists and locals when computing the initial score for each station. One way to compute the score for a station is using the probability of being a tourist, given that a commuter has visited that station. For simplicity, we simply denote this probtourist andability.

5.5 TOURIST PREFERENCE ANALYTICS SYSTEM DESIGN

After the tourists are identified from the public commuters, their travel information, especially their travel locations can be directly obtained from their transport riding records. Based on such information, the system conducts the analytics for tourist location preferences. Specifically, our current design is to predict and recommend (1) the next public transport alighting location, i.e., the corresponding

attraction that a tourist will visit,2 and (2) the associated next public transport boarding location, i.e., the end point of the tour. The first one provides the basic personalized recommendation service, while the second one enhances the personalization and the comprehensiveness of the services.

Model Description

We denote a set of the identified tourists by U ¼ fu$_1$;u$_2$;:::;ujU j g and a set of locations (i.e., subway stations and bus stops) by L ¼fl$_1$;l$_2$;:::;l$_j$L$_j$ g. Note that l$_i$ 2 L contains the descriptive information such as station name, longitude, latitude, etc. For each tourist u 2 U, her historical transport records, i.e., tours (in chronological order) is denoted by C$_u$ ¼fc$_1$;c$_2$;:::;c$_n$}, where each tour c$_i$ ¼< l$_x$;l$_y$ > consists of a public transport alighting location l$_x$ and the next boarding location l$_y$. We first build the tourist-location visit frequency matrix M$_{ul}$ to record the count of visits between tourists and locations. Intuitively, the higher the visit frequency is, the more the location is preferred by the corresponding tourist. In order to estimate a tourist's ranking preference for locations, we first denote tourist u's preference for location l by a binary variable '$_{u;l}$: 'u;l ¼10 ifif rru;lu;l >¼ 00

5.6 TENSORFLOW

Currently, the most famous deep learning library in the world is Google's TensorFlow. Google product uses machine learning in all of its products to improve the search engine, translation, image captioning or recommendations. To give a concrete example, Google users can experience a faster and more refined the search with AI. If the user types a keyword a the search bar, Google provides a recommendation about what could be the next word. Google wants to use machine learning to take advantage of their massive datasets to give users the best experience. Three different groups use machine learning:

- Researchers
- Data scientists
- Programmers.

They can all use the same toolset to collaborate with each other and improve their efficiency. Google does not just have any data; they have the world's most massive computer, so Tensor Flow was built to scale. TensorFlow is a library developed by the Google Brain Team to accelerate machine learning and deep neural network research.

It was built to run on multiple CPUs or GPUs and even mobile operating systems, and it has several wrappers in several languages like Python, C++ or Java.

TensorFlow Architecture

Tensorflow architecture works in three parts:

- Preprocessing the data
- Build the model
- Train and estimate the model

It is called Tensorflow because it takes input as a multi-dimensional array, also known as tensors. You can construct a sort of flowchart of operations (called a Graph) that you want to perform on that input. The input goes in at one end, and then it flows through this system of multiple operations and comes out the other end as output. This is why it is called TensorFlow because the tensor goes in it flows through a list of operations, and then it comes out the other side.

5.7 RESNET 50 MODEL IMPLEMENTATION

ResNet-50 is a convolutional neural network that is 50 layers deep. You can load a pretrained version of the network trained on more than a million images from the ImageNet database [1]. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224.

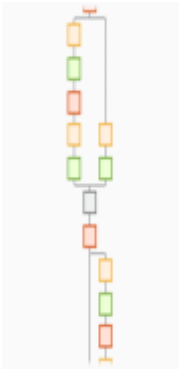

Fig. 3: ResNet50

You can use classify to classify new images using the ResNet-50 model. Follow the steps of Classify Image Using GoogLeNet and replace GoogLeNet with ResNet-50.To retrain the network on a new classification task, follow the steps of Train Deep Learning Network to Classify New Images and load ResNet-50 instead of GoogLeNet.

net = resnet50 returns a ResNet-50 network trained on the ImageNet data set.This function requires the Deep Learning Toolbox™ Model for ResNet-50 Network support package. If this support package is not installed, then the function provides a download link.net = resnet50('Weights','imagenet') returns a ResNet-50 network trained on the ImageNet data set. This syntax is equivalent to net = resnet50

5.4. SOFTWARE REQUIREMENTS

Operating System: Windows 7
User Interface : HTML, CSS
Client-side Scripting: JavaScript
Programming Language : Python
Web Applications: JDBC, Servlets, JSP
IDE/Workbench: My Eclipse 8.6
Database: Oracle 11g
Server Deployment: Tomcat 7.0

5.5. HARDWARE REQUIREMENTS

Processor :Intel core i3 or above
Hard Disk :500GB

6. RESULTS

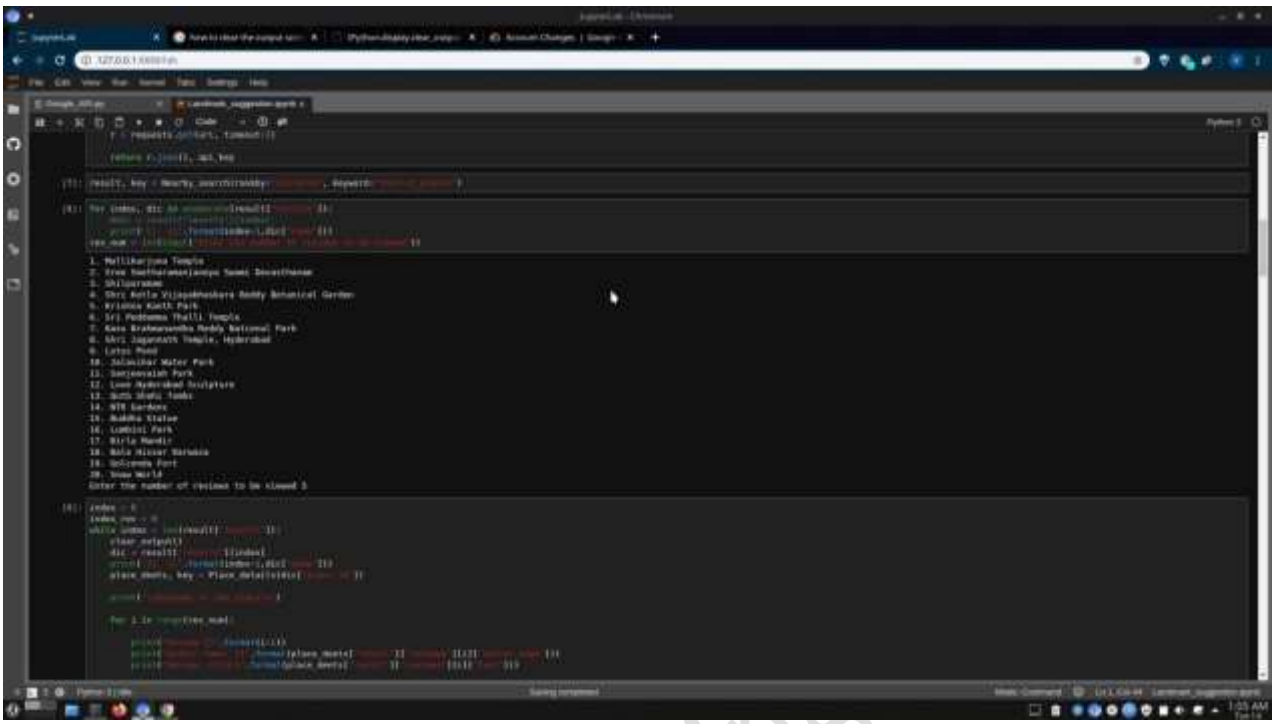

Fig. 4 List of Tourist Places

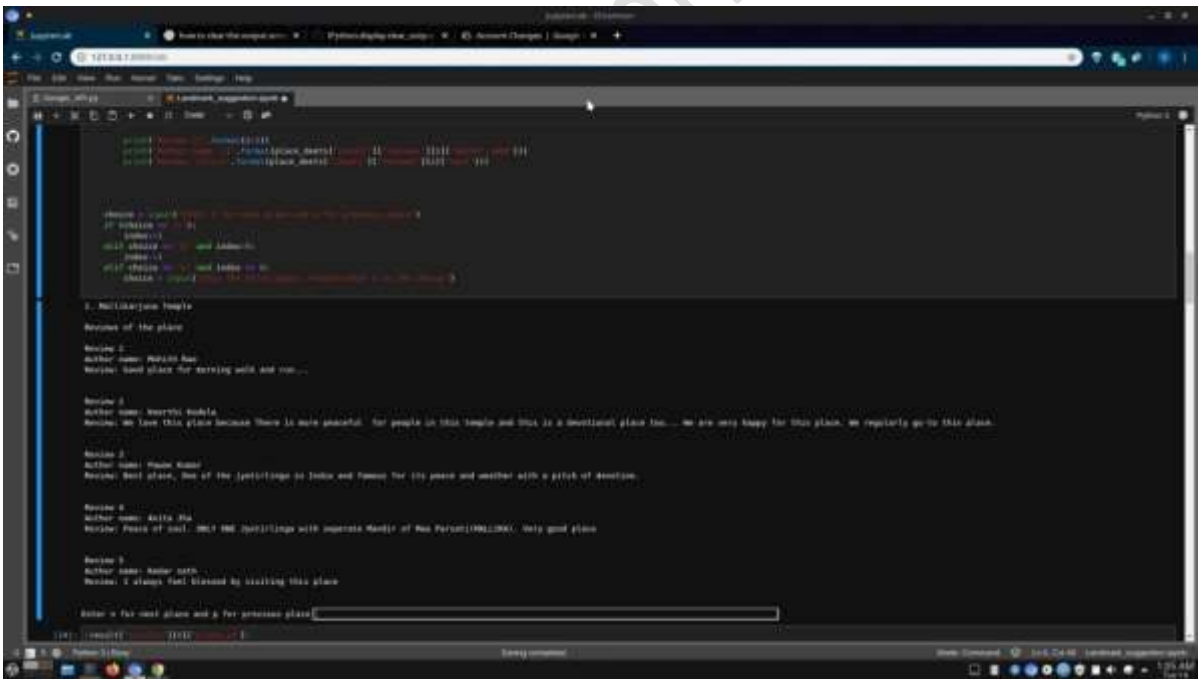

Fig. 5 Reviews about the selected tourist place

7. CONCLUSION

Therefore, through this paper we try to achieve an efficient and hassle free approach towards finding information on the monuments and finding routes to the preferred monuments. This program can also be used to find routes to a place with minimal approach roads.

For future expansion we plan to include many other features like, finding eateries nearby, and adding them to the route of

the monument, adding places which are meant to be for site-seeing rather than visiting while en route to the preferred monument and many other such features.

REFERENCES

[1] [1]. Lu, Y., Wu, H., Liu, X., & Chen, P. (2019). TourSense: A Framework for Tourist Identification and Analytics Using Transport Data. IEEE Transactions on Knowledge and Data Engineering, 31(12), 2407–2422. doi: 10.1109/tkde.2019.2894131

[2] Al Rabbani Alif, Mujadded & Ahmed, Sabbir & Hasan, Muhammad. (2017). Isolated Bangla handwritten character recognition with convolutional neural network. 1-6. 10.1109/ ICCITECHN.2017.8281823.

[3] Napoletano, Paolo & Piccoli, Flavio & Schettini, Raimondo. (2018). Anomaly Detection in Nanofibrous Materials by CNN-Based Self-Similarity. Sensors (Basel, Switzerland). 18. 10.3390/ s18010209.

[4] Kim, Jaebok & Englebienne, Gwenn & Truong, Khiet & Evers, Vanessa. (2017). Deep Temporal Models using Identity Skip- Connections for Speech Emotion Recognition. 10.1145/3123266.3123353.

[5] [5]. Real Python. (2020, January 23). The Ultimate Guide to Speech Recognition with Python.

[6] Retrieved from https://realpython.com/python-speech-recognition/

[7] [6]. Keras overview: TensorFlow Core. (n.d.). Retrieved from https://www.tensorflow.org/guide/ keras/overview

[8] [7] Fung, V. (2017, July 17). An Overview of ResNet and its Variants. Retrieved from https:// towardsdatascience.com/an-overview-of- resnet-and-its-variants-5281e2f56035..