In the world as we know it, hand-held computing devices have penetrated our lives. This is tolerated mostly due to the fact that many of the technologies that these devices provide create a comforting level of convenience. In fact, society expects more and more from these small, life augmenting devices. The simple cell phone transformed from just a “phone" into a powerful “smart device" that today has as much computing power as a desktop computer from a decade ago. In fact, a smart device offers quite a bit more functionality than a decade old desktop in a remarkably smaller form. Aside from the common uses (phone, camera, video recorder), “smart devices” are even more powerful with integrated chip sets such as global positioning systems (GPS) to measure geospatial location and accelerometers that can measure a devices orientation.

With rapid growth in the technological industry, mobile devices become an essential item of people’s lives due to the fact that these devices provide a lot of features which helps people to do their routine activities in much easier way. Therefore society expects more from these devices which leads device vendors & application developers to do more R&D and enhance the features. Due to this, mobile devices which were having the functionality of just making a call became devices which can perform “smart activities”. They have become more powerful with the integration of chip sets such as global positioning systems (GPS) to measure geospatial location and accelerometers that can measure a devices orientation.

Even with the popularity of GPS as a tool for spatial data collection, in some researches it is claimed that GPS has poor user-time coverage and long gaps because satellites are cut off indoors or under cover where many people spend the vast majority of their time. And the average time associated with GPS coverage was a mere 4.5%, whereas GSM and 802.11 coverage were 99.6% and 94.5%, respectively. [LaMarca, et. al., 2005]

Even though the popularity of the GPS is high, in some researches it has claimed that using GPS as method to collect user mobility data and geographical data is not much accurate since GPS data may contain long gaps & poor user time coverage because signal establishment with the satellites are cut off in indoors or due to the impenetrable covers where people stay most of their time. And the average time associated with GPS coverage was a mere 4.5%, whereas GSM and 802.11 coverage were 99.6% and 94.5%, respectively. [LaMarca, et. al., 2005]

In actuality, GPS receivers have no problem receiving signal as long as the receiver has at least 4 partial view of the sky. The GPS satellite constellation consists of 24 satellites (plus some spares) orbiting in six different planes. Each of these planes are inclined 55◦ from the Earth’s equatorial plane. The satellites are positioned in their respective planes in such a manner that from almost any place on Earth, at least four satellites are above the horizon at any time [Ashby N., 2003]. With this configuration, every receiver is nearly always guaranteed to be in view of the minimum number of four satellites needed to get an accurate fix (assuming there are no obstructions). If enough satellites are in view, an accuracy within two meters can be achieved (5-10 meters is a realistic expectation [Wolf, 2006]). This spatial accuracy coupled with GPS satellites’ extremely accurate clocks allows for great representation of a user’s mobility.

The spatial accuracy coupled with the high accuracy clocks of GPS satellites allows for a great representation of a user’s mobility. There are 24 satellites (plus some additional satellites) orbiting earth in six different planes. Each of these planes are inclined 55◦ from the Earth’s equatorial plane. These satellites are positioned in their respective planes in way that at least four of them are above the horizon every time from almost any place on the Earth. As long as GPS devices are having at least 4 partial view of the sky, they are not having any problems of receiving signals [Ashby N., 2003]. By assuming that there are no obstructions, to get an accurate fix by GPS receivers at any given time they are nearly always guaranteed to be in view of the minimum number of 4 satellites. If enough satellites are in view, an accuracy within two meters can be achieved (5-10 meters is a realistic expectation [Wolf, 2006]).

Processing GPS data encompasses the following: filtering, smoothing and interpolation. Processing is defined as repairing or putting through a prescribed procedure". Each feature individually performs an essential task that ultimately determines unfavorable attributes and either identifies them or removes them. One prime reason for the processing of a collected set of GPS streams is to replace the impossible task of visually inspecting the collection [Jun J., 2005].

The spatial data collected using GPS need to be processed before using. Processing is defined as repairing or putting through a prescribed procedure which contains the steps of filtering, smoothing and interpolation. The main reason for processing this collected data is to replace the impossible task of visually inspecting the collection. When processing each step performs an essential task which determines unfavorable attributes and either identifies or removes them [Jun J., 2005].

According to [Kardashayn, 2011], due to the complicated traffic networks, traffic speed and the huge number of the traffic participants, the safety cameras and other existing traffic management methods are not good enough for controlling and managing traffic in any situation and in any location. [Kardashayn, 2011] describes a new traffic management solution based on the automatically individual control to any traffic user anywhere and anytime. The principle of the method is as follows: any registered vehicle periodically sends information about itself, which is being decoded and analyzed by the central traffic management unit. As a result the central traffic management unit knows the location, speed and condition for every single registered vehicle. The system can establish traffic management due to the traffic management algorithm. [Yoon, et. al., 2007] proposes a simple yet very effective method that can capture traffic states in complex urban areas. For evaluation, they applied their system to two different GPS trace data sets collected in the Ann Arbor in Michigan. The results showed that accuracy of higher than 90% can be achieved if ten or more traversal traces are collected on each road. Moreover, traffic patterns turned out to be fairly consistent over time, which allowed the use a larger history in classifying traffic conditions. [Thianniwet, et. al., 2009], proposed a technique to identify road traffic congestion levels from velocity of mobile sensors with high accuracy and consistent with motorists’ judgments. The data collection utilized a GPS device, a webcam, and an opinion survey. Human perceptions were used to rate the traffic congestion levels into three levels: light, heavy, and jam. The ratings and velocity were fed into a decision tree learning model. They successfully extracted vehicle movement patterns to feed into the learning model using a sliding windows technique. The model achieved accuracy as high as 91.29%. Biem et al. [Biem, et. al., 2010] describes some of their recent work in supporting real-time Traffic Information Management using a stream computing approach. They used GPS data from some taxis and trucks to highlight some of their findings on traffic variability in the city of Stockholm. Their customized analyses include continuously updated speed and traffic flow measurements for all the different streets in a city, traffic volume measurements by region, estimates of travel times between different points of the city, stochastic shortest path routes based on current traffic conditions, etc. In order to benefit from telematics based data collection, time-dependent travel time estimates have to be integrated into time dependent vehicle routing frameworks. [Ehmke & Mattfeld, 2010] discusses data collection and the conversion from raw empirical traffic data into information models, an application example compare several information models based on real traffic data regarding its benefits for time-dependent route planning. The integration of information models into time dependent vehicle routing frameworks is discussed. The data mining approach as in [Ehmke & Mattfeld, 2010] provides time-dependent travel times in a memory efficient way without a significant reduction of the itineraries’ reliability and robustness. [Tripathi, 2010] presents an algorithm for detection of hot spots of traffic through analysis of GPS data by analyzing two data clustering algorithms: the K-Means Clustering, and the Fuzzy C-Means Clustering. After the clustering process stops, a cluster center can be selected, which will display the membership grades of all data points toward the selected cluster center. They justify the fact that they use clustering algorithm for the detection of the hot-spots, where each cluster represents the group of GPS data points having latitude and longitude as their coordinate and very small distance between them. To measure the distance between two points on the earth surface, [Tripathi, 2010] derived a formula for calculating geodesic distance between a pair of latitude/longitude points on the earth‘s surface, using the WGS-84 ellipsoidal. A move to try to understand, manage and predict the traffic phenomenon in a city is both interesting and useful.

Throughout this research it has been proven that to determine a best transportation method in Sri Lankan context, the data mining concepts together with crowdsourcing can be applied. And based on a thorough analysis by extending the data set of the collection stage, this research can be extended to predict the best transportation method in all the areas in Sri Lanka.