An adaptive transportation prediction model for the informal public transport sector in Africa

I. Ndibatya[1], M.J. Booysen[2] and J. Quinn[3]

Abstract - **The informal public transport sector in Sub-Saharan Africa is responsible for transporting the overwhelming majority of the workforce. Often, passengers have to wait for hours for taxis to coincidentally pass by to pick them up, making the transport mode notoriously inefficient. Despite its relevance and impact, the sector is afforded little attention in terms of regulation, development and organization, giving rise to a complex and inefficient system that affects millions of people. In fact, little is known about the industry. To advance understanding of this system, minibus taxis were equipped with tracking devices in this study. Tracking data was then used to develop a model that describes the transport network – essentially finding patterns in the apparent chaos for the potential benefit of its users. The adaptive model uses unsupervised learning to predict the informal stages in the city and provide travelers with intelligence on the best time and place to get transport, thereby reducing the waiting time at the taxi rank and the informal roadside stops.**

Experimental results show 70.4% model accuracy in dynamically learning the taxi behavior and accurately predicting the best places to get taxis at a given time of the day.

Index Terms: Adaptive Model, Informal Transport Sector, Transport Prediction, Intelligent Transport Systems.

I. Introduction

Road transport accounts for 90% of passenger traffic on the African continent. South Africa, in particular has the largest distribution of the road network, with a road to population ratio of 56.3km for every 10,000 population on the continent [1]. In sub-Saharan Africa various strides have been made toward the implementation of Intelligent Transport Systems (ITS) such as: Freeway Management System (FMS) in Cape Town and Durban; Bus Rapid Transit programs in Johannesburg, Durban, Cape Town and etc. [2]. However, these programs provide little coverage in the overall public transport sector compared to taxis.

A. *The informal public transport sector*

Urbanization in Africa has recently increased to an estimated 33%, and is expected to increase to 50% by 2020. This growth has been a strain on the available services such as water, electricity, transport and others [3]. The strain on the transport services is compounded by a lack of planning and investment, which has given rise to an on-demand alternative transport sector, the informal public transport sector (*IPTS*)[3].

The IPTS has consistently grown unplanned and uncontrolled since early 1990s after the collapse of the governments' bus transit systems in East, Central and Sub-Saharan Africa [3]. Studies on the current state of IPTS are limited, but state that the small-sized mini-buses that carry 8-25 passengers represent the lion’s share in the industry [3]. These mini-busses are called different names, such as *trotro* in Ghana, *danfo* in Nigeria, *gbaka* in Ivory Coast, *matatu* in Kenya and *taxi* in Uganda. According to [2], Lagos alone had 80,000 minibuses and South Africa had more than 200,000 minibuses in 2012.

Minibuses have filled the transportation gap, especially for the poor people in developing countries. However, they present clear disadvantages from the perspective of public interest. These include [2, 3]:

- *Road congestion:* Minibuses now account for more than 50% of motorized traffic on the road. Although they give transport to the masses, the unpredicted stopping and reckless driving has a considerable negative effect on the traffic flow in urban areas
- *Safety and emissions.* Most minibuses are old, inadequately maintained and are operated for long hours. They are responsible for more than 20% of the road accidents and significantly contribute to the carbon emissions.
  - *Unpredictability of routes, schedules and fares.* Though governments have tried to regulate minibuses by building taxi ranks and designating taxi stops, the behavior of the drivers remain unpredictable in terms of route selection, schedules and fares. Sometimes the routes, schedules and fares depend on the availability of passengers. The stops along the route are unpredictable, and normally requested on an ad-hoc basis.

These areas of concern open opportunities for the novel application of intelligent transport systems (ITS) to make the informal public transport sector safer, more efficient and greener.

Normally, passengers don't know the exact mobility patterns of the taxis. To get a taxi, the passengers have to either go to the taxi rank, or wait by the road side at a tacitly known taxi stop. In the taxi rank, the passengers sit in the taxi and wait for it to get full before departure. At the roadside stops, the passengers have to wait for an unpredictable period of time to get a taxi. The waiting time is normally between 15 minutes to 2.25 hours. For a sector that transports over 77% of the workforce, a waiting time in excess of an hour aggregates to a considerable overall time

[1] I. Ndibatya is an exchange student with the Department of Electrical and Electronics Engineering, Stellenbosch University, South Africa. indibatya@gmail.com

[2] M.J. Booysen is a Lecturer at the Department of Electrical and electronics Engineering, Stellenbosch University, South Africa. mjbooysen@sun.ac.za

[3] J. Quinn is a Senior Lecturer with the department of Computer Science, Makerere University, Uganda. jquinn@cis.mak.ac.ug

Preprint submitted to 17th International IEEE Conference on Intelligent Transportation Systems. Received June 21, 2014.

loss, which has a stunning negative effect on the economy and other related outcomes.

B. Contribution

This paper presents an adaptive model to solve the efficiency problem in the town-service minibus taxis. The model is adaptive because it makes use of the tracking data to learn the behavior of the drivers and update its prediction methods accordingly.

Minibus taxis were equipped with GPS-enabled fleet management devices that continually report geo-location information. We applied Machine Learning (Unsupervised Learning) Techniques to the data to develop a model for the movement patterns of the taxis. Specifically, the model predicts the likelihood of a passenger being picked up from a given location, at a given time and a given day of the week. The model was validated against 30% of the recorded data, which was not used for the development of the model. In addition to predicting the best places to wait for the taxi, it also gives alternative places for slightly higher waiting time, depending on the location of the passenger.

The rest of this paper is organized as follows: Section II provides the scientific basis for the research or Literature review; Section III describes the research design and model development methods used ; Section IV model testing and results ; Section V the discussion of results; Section VI concludes the paper. In this paper we also use the term cluster and stops interchangeably to mean the organic taxi stops discovered during the research.

II. Scientific Basis for the Research

We based our research on a recent approach that has received great attention due to the emergence of powerful computing – Data driven modeling (DDM) [4]. DDM is based on analyzing data about a system and finding relationships between variables without explicitly having knowledge of the behavior of the system. Figure 1 shows the main process of data driven modelling.

DDM has successfully been used in the fields of life sciences, natural sciences, biological sciences, Earth Sciences [5], finance and computational sciences to uncover hidden relationships in vast amounts of data. DDM combines techniques from various fields which include Artificial Intelligence, Data-mining, Intelligent Data Analysis and Machine Learning.

Studies have been done in developed countries related to the transport services with random routes such as: Many road shipments, pick-up and delivery services, express couriers, demand responsive public services, and public utility services (gas, water, electricity) [6]. DDM techniques have been applied during these studies and have proven to be useful. Automatic Vehicle Location Systems (AVLS) have been deployed in vehicles to pinpoint their locations and assign them the next available task. In such services, the average service time *T* has been generalized as a function of: the number of operating mobile units in the fleet, *N*; the vehicle operative area *S*; and the average speed of the mobile unit *v*.

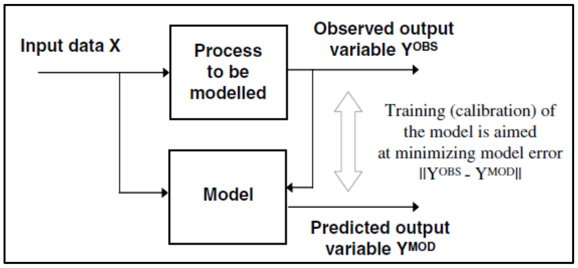

Input data X
Process to be modelled
Observed output variable Y^{OBS}
Training (calibration) of the model is aimed at minimizing model error $||Y^{OBS} - Y^{MOD}||$
Model
Predicted output variable Y^{MOD}

Fig 1. Figure showing the process of data driven modelling. [4]

Another related field of research is in Demand Responsive Transit (DRT) systems in the USA where passengers call in to the providers to be picked up from one location to the other. In these systems, providers are given a time window within which to pick up the passengers. Beyond this time window, the providers are fined. DRTs have worked well with the elderly and the disabled. These systems are greatly subsidized by government. It has been established that the cost of running the DRT increases with the reducing of the time window because operators have to increase the size of their fleet to beat the time constraint [7]. Still in these studies, DDM techniques have been used and the results have been positive.

The Informal Public Transport Sector (IPTS) in Africa operates some form of organic Demand Responsive Transit (oDRT) system depending much on the behavior of the driver and the demand. For example, a taxi has to get full before it leaves the taxi rank. On the route to destination it stops to off-load and load passengers. (Note that the stops are informal and dynamic depending on the destinations and availability of passengers). There is no defined window to wait for the taxi and there are many informal stops on the way to the destination. The waiting time at these stops is also not known but ranges between 15 minutes to 2.25 hours.

It is on this basis that we sought to study the behavior of minibus taxis in the informal public transport sector of South Africa with a view of applying DDM techniques (machine learning in particular) to develop an appropriate ITS application to address some of the challenges facing the Informal Public Transport Sector (IPTS).

III. Research Design and Model Development

A. Data collection

Fleet management units were installed in ten minibus taxis that operate between Stellenbosch and Somerset West from November 2013 to May 2014. Geo-location data was collected for six months at a nominal frequency of 1 Hz. The data logged include: date and time, latitude, longitude, speed and direction. By the end of the six-month period, a total of 1,185,261 GPS locations from the ten minibus taxis had been collected.

B. Data classification

Preliminary spatial analysis of the data illustrated in Figure 2 showed six areas of data concentration i.e. Stellenbosch, Eden Town, James Town, Techno Park, Somerset West and Strand. The other sparse concentration is on the main road that connects the six towns. To achieve more focused results, the study area was divided into six zones according to the outcomes of the preliminary analysis.

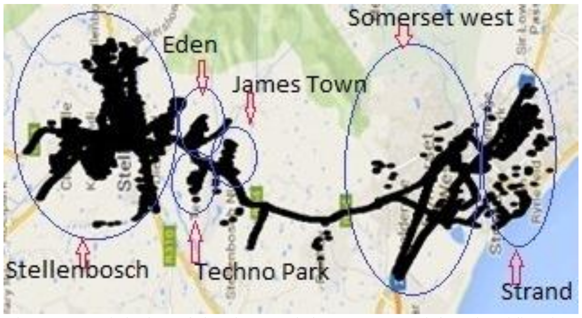


Fig. 2: Spatial plot of the data and the zones used for classification.

A classification algorithm was developed (Figure 3) to classify the data according to the six zones.

The classification algorithm classifies the observations according to Euclidian distance between the central point (Base Longitude and Latitude) and the observation. If the distance is less than the defined zone radius, then the observation belongs to the current zone. Table I shows the Zones, central points and radii used to classify the data.

The algorithm was implemented in R programming language [9]. It runs well on a single core processor with a time complexity of O(n) for a constant classification set and O(n^2) for a constant tracking geodata set. When run on a multicore computer with *k* cores and clustering enabled, the time complexity is O(n)/k for a constant classification set and O(n^2)/k for a constant geodata tracking set.

TABLE I
CLASSIFICATION ZONES, RADIUS and BASE GEO-POINTS USED FOR CLASSIFICATION.

No.	Zone	Radius (km)	Latitude	Longitude
1	Stellenbosch	4	-33.921872	18.855336
2	Eden	1	-33.964570	18.857449
3	Jamestown	1.5	-33.980087	18.835820
4	Techno Park	1	-33.965424	18.838052
5	Somerset-west	4	-34.065879	18.841485
6	Strand	3	-34.111006	18.848424

```
read $D_{tr}$  #Geodata csv
read $D_{cl}$  #Classification csv
initialize RD #Data frame to hold the results
for each item in $D_{cl}$ ($D_i$)
    get radius $R_i$ # reference radius
    get coordinate C($lat_i$,$long_i$) #the reference geo coordinate
    for each item in Dtr ($T_j$)
        get coordinate X($lat_j$,$long_j$)  #Tracker geo coordinate
        Ed <-- dist(C,X)    #Euclidean distance
        if (ED<=$R_i$) then
            #append classifying info  to the tracking geodata
            $RD_j$ <--append($D_i$,$T_j$)
        end if
    end for
end for
Unique (RD,j)  #subset RD by unique j
return RD # data-frame of classified tracking geodata
```

Fig. 3: Figure showing the data classification algorithm.

C. Model development

We analyzed the data using unsupervised learning techniques of machine learning, i.e. clustering. A density-based clustering algorithm –DBSCAN was used on the data for Stellenbosch. Density Based Spatial Clustering of Applications with Noise (DBSCAN) can discover patterns of any shapes in large data sets [8]. The DBSCAN algorithm classifies data with respect to the Radius (*Eps*) and the minimum points in the cluster (*MinPts*)

i) Description of the sample

The spatial plot in Figure 2 indicates a lot of activity in Stellenbosch. The classification results gave 1,074,050 tracking positions (90.6% of the total observations) located in Stellenbosch alone. The following investigation was done for the Stellenbosch data only. To investigate the organic stops by the taxis in Stellenbosch, only data for which the speeds were less than 1km/h was selected. This criteria means that only data recorded when the taxis were either going to stop, had stopped, or were departing from the stops were selected. A total of 81,525 data points were obtained from this selection. Due to performance limitations with a very large data set and the computing resources we had during the research, we randomly selected 70% of recorded Stellenbosch observations where the speeds were less than 1km/h, giving a final sample of 57,068 observations for model formulation and training. We then used the remaining 30% (24,457) for model testing.

ii) Clustering

The DBSCAN algorithm was run on the sample of 57,068 observations with *Eps=0.0002* and *MinPts=70*, resulting in 57 clusters of varying densities. Figures 4 (a and b) show the spatial plots of the dataset before and after clustering.

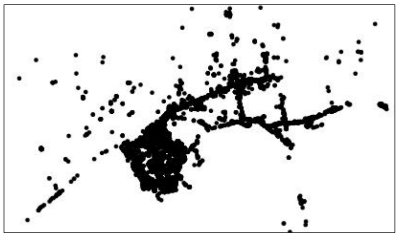


Figure 4(a): Spatial plot of sample data before clustering

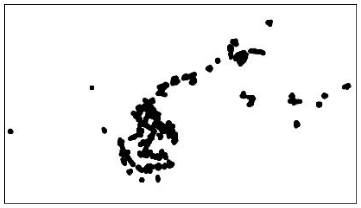


Figure 4(b): Spatial plot of sample data after clustering

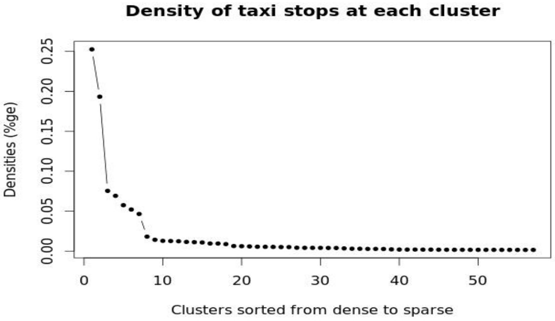


Fig. 5: Figure showing the variation of stop densities with clusters.

It is clear in plot 4(b) that the algorithm eliminated outliers and maintained the clusters at dense points that are most likely the documented and undocumented dynamic stops where the taxis pick up and drop passengers. We further analyzed the densities of the different clusters. Figure 5 shows the clusters sorted with decreasing densities.

We further analyzed activity at individual clusters to discover whether there are patterns according to different days of the week and the different times of the day. We sampled seven clusters from the clusters produced by the DBSCAN algorithm and subjected them to this analysis.

The results in Figure 6 show that there are different patterns at the different selected clusters on the different days of the week. Every line in the graph indicates one cluster and how the densities vary during the different days of the week.

Finally we analyzed activity at the individual clusters for the different times of the day divided into 30 minutes time segments. Figures 7(a, b and c) show that the patterns at different clusters vary for every day and time of the week. Figure 7a shows a clear peak during morning hours for cluster 2, and Figure 7b shows increased activity for cluster 1 in the afternoons. After evaluating the geo-locations of these clusters, it was found that cluster 1 is the main taxi rank in town (where people head to work in the mornings), and that cluster 2 is the main taxi rank in the urban dwelling (Kayamandi). Cluster 3 (Figure 7c) shows a distribution that is less time-dependent. This will be discussed further in section V

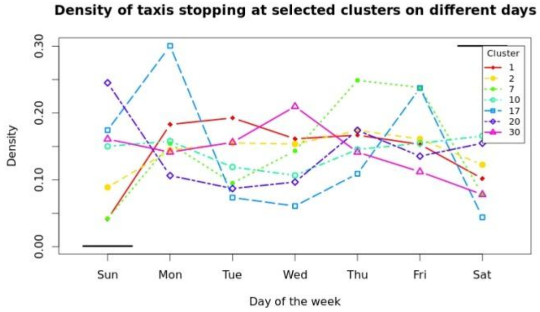


Fig.6: Figure showing the varying densities of the taxis stopping at the identified clusters on the different days of the week.

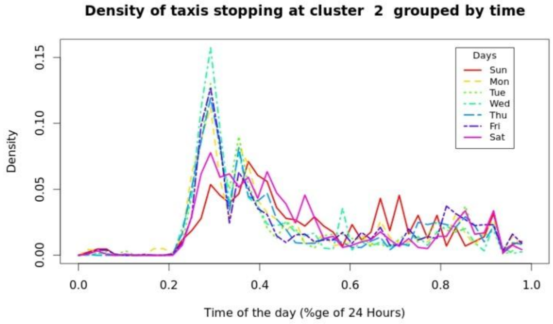


Fig. 7 (a): Figure showing the variation of densities of taxis stopping at cluster 2 on different days.

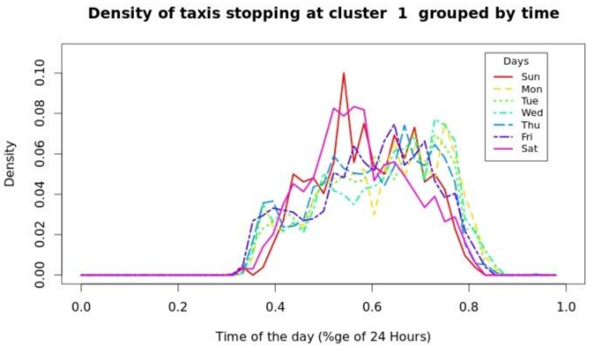


Fig. 7 (b): Figure showing the variation of densities of taxis stopping at cluster 1 on different days.

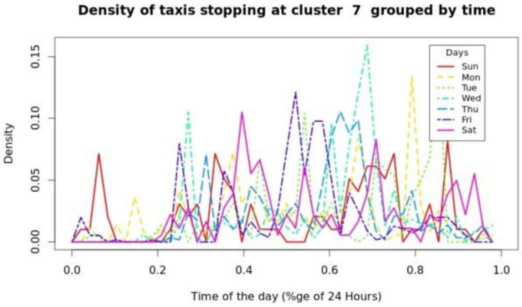


Fig. 7 (c): Figure showing the variation of densities of taxis stopping at cluster 7 on different days.

iii) The adaptive transportation location model

From part (ii) above, it was discovered that locating a taxi in Stellenbosch depends on, where you are (*L*), the day of the week (*D*) the time of the day (*T*).

An adaptive transport prediction model (Figure 8) was developed by performing probabilistic analysis for every parameter *(L, D & T*) in the training data set as shown in figure 8. When the model runs, the output is a model matrix (Sample demonstrated in tables II & III).

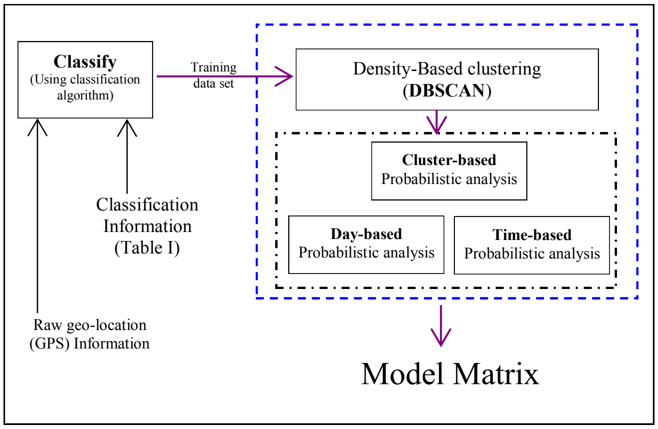

Classify
(Using classification algorithm)
Training data set
Density-Based clustering (DBSCAN)
Cluster-based Probabilistic analysis
Day-based Probabilistic analysis
Time-based Probabilistic analysis
Classification Information (Table I)
Raw geo-location (GPS) Information
Model Matrix

Figure 8: figure describing the adaptive transport location model.

TABLE II

PARTIAL MODEL MATRIX TABLE SHOWING THE PROBABILITIES WITH RESPECT TO DAY AND TIME SEGMENTS

Day	**Probability in percentage of stopping at cluster 4 for each time segment during the week**																
	8:00	**8:30**	**9:00**	**9:30**	**10:00**	**10:30**	**11:00**	**11:30**	**12:00**	**12:30**	**13:00**	**13:30**	**14:00**	**14:30**	**15:00**	**15:30**	**16:00**
Sun	13	20	22	0	6	28	0	16	10	25	6	4	27	4	18	20	21
Mon	7	0	26	0	0	11	0	8	10	0	0	4	7	0	11	0	0
Tue	0	0	4	12	6	0	60	14	10	0	18	0	0	8	4	0	3
Wed	0	20	26	12	0	0	0	3	0	0	0	9	7	0	0	16	8
Thu	7	33	9	29	19	17	13	0	0	0	0	61	27	13	4	0	5
Fri	33	13	4	24	31	17	13	54	70	50	24	4	0	75	61	20	39
Sat	40	13	9	24	38	28	13	5	0	25	53	17	33	0	4	44	24

TABLE III

PARTIAL MODEL MATRIX TABLE SHOWING THE PROBABILITIES WITH RESPECT TO CLUSTER AND THE DAY OF THE WEEK

Cluster stops		**%ge Probability of stopping at a cluster**						
Cluster	**Proportion**	**Sun**	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**
1	25	11	29	31	27	24	24	26
2	19	18	17	19	20	19	20	24
3	8	12	6	8	7	8	7	6
4	7	11	6	7	6	7	5	9
5	6	7	6	6	4	6	8	4
6	5	7	5	4	5	5	6	5
7	5	2	4	3	4	7	7	4

The model matrix is a collection of all possible probabilities of locating a taxi in Stellenbosch at any of the clusters (dynamic stops) identified in part (ii). The matrix is generated using the training data set described in section C subsection (i) and stores the probabilities in three dimensions.

IV. Model Testing and Results

To test the model we used another unsupervised learning technique (K-means clustering) to cluster the test data according to the cluster centers deduced in part III. K-means clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. We ran the k-means algorithm with k=58 and n=24,457. This method was used because at this point we already knew the clusters.

The k-means clusters formed were then input into the model testing component as illustrated by Figure 9.

During the testing phase, the assumption was made that the following does not impact on the taxi distributions.

- The weather conditions
- The season and time of the year
- The traffic flows and patterns

Other implicit assumptions are:

- A taxi stopped to either drop or pick up a passenger from the identified cluster/stop, and that there is space in the taxi.
- The 10 tracked taxis are the only taxis taking passengers. The accuracy would increase if more taxis are sampled.

A. *Model predictions and actual observation*

We tested the adaptive transport location model on a set of 192 scenarios which are a combination of Location of the passenger (L), day of the week (D) and time of the day (T)

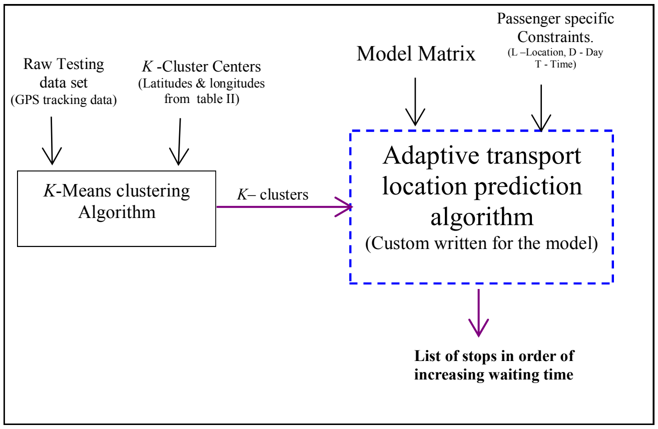
Four locations (L=4), Three days (D=3) and sixteen time segments (T=16) were used in the scenario set. Table IV shows the sample outcome of the model predictions. The predictions are made based on the model matrix generated in section III. To make a prediction, the model scans through all the available stops and finds the five closest stops to the location of the passenger. Then it computes the probability of finding a taxi at each of the stops with respect to the day and time. Finally it ranks the stops according to probability. I.e., the stop with highest probability ranks 1 & so on as shown in Table IV.

We used the test set clustered using the k-mean algorithm to get the actual observations. We selected only data for the five nearest stops to the location of the passenger and we performed statistical analysis on the data. The output from this exercise was a set of clusters ranked according to actual observed activity at the cluster for each of the 192 test scenario. Table IV shows a partial comparison of the model predictions with the real output. (Note: In table IV, items preceded with a *p* are predictions where as those preceded with *a* are actual.)

With the 192 test cases, the model achieved 70.4% accuracy for the best stops.

V. Discussion

The initial analysis of the data in section III revealed that taxis in Stellenbosch like in many other cities in Sub-Saharan Africa do not stop in designated places only. They form a web of organic stops (figure 4b) whole over the route. However, these stops are not completely random. This is because when we run the DBSCAN clustering algorithm 10 times on the same dataset, the number of clusters formed ranged between 57 to 60 and there was persistence of 90% mainly among the large clusters. It is yet to be established whether the positions of these clusters change over a given period of time.



Raw Testing data set (GPS tracking data)
K -Cluster Centers (Latitudes & longitudes from table II)
Model Matrix
Passenger specific Constraints. (L –Location, D - Day T - Time)
K-Means clustering Algorithm
K– clusters
Adaptive transport location prediction algorithm (Custom written for the model)
List of stops in order of increasing waiting time

Figure 9: Figure showing the model testing phase

TABLE IV
PARTIAL TABLE OF MODEL PREDICTIONS AND ACTUAL OBSERVATIONS IN ORDER OF INCREASING WAITING TIME

Reference Location	-33.9, 18.8				
Nearby clusters	30	54	44	33	50
Distance from ref(Km)	1.94	1.97	2.05	2.07	2.08

Results

	rank	1	2	3	4	5
1	**p.Mon.15:30**	**54**	**30**	**44**	**33**	**50**
	a.Mon.15:30	54	33	44	30	50
2	**p.Mon.21:30**	**30**	**54**	**44**	**33**	**50**
	a.Mon.21:30	30	33	44	54	--
3	**p.Thu.11:00**	**50**	**30**	**54**	**44**	**33**
	a.Thu.11:00	50	33	44	54	--
4	**p.Sat.15:30**	**30**	**54**	**44**	**33**	**50**
	a.Sat.15:30	54	33	44	30	50

In addition, the number of taxis stopping at any cluster varies with the day of the week (figure 6) and the time of the day. Figures 7(a, b & c) show that at cluster 2, most taxis stop very early at around 7:00 am while at cluster 1 there is a close to normal distribution between 9:00 am and 8:00 pm with the mean around mid-day. While at cluster 7, there is an almost uniform activity for most of the day. Several reasons could explain this behavior. 1) Stops with a lot of activity in the morning could be the stages within the working neighborhoods that go to work every morning. 2) Stops with a normal distribution between 10:00 am and 8:00 pm could be town stops always active during the day 3) stops with uniform activity throughout the day and night could be the random organic stops by the road sides.

This information could be useful for transportation systems planners to optimize resource allocation. Passengers looking for taxis could use the information to plan their journey in advance. This is because of the intelligence on where to find taxis given to them by the developed ITS hence saving on the time wasted during the whole process.

VI. Conclusion

We have developed a transport prediction model that can adapt to the ever-changing behavior of the taxi drivers in the informal public transport sector. Since the inefficiency problem of the IPTS is common to most cities in Sub-Saharan Africa and Africa in general, the model can be useful for all developing countries if trained using the data collected from that country.

With more data spanning over a year and more computing resources to train the model, more training parameters can be incorporated in the model such as weather conditions, directions and season. We believe that adoption and improvement of this model for the informal public transport sector could significantly improve the efficiency of the sector.

VII. Further Research

Many interesting research problems in ITS applications to the informal public transport sector remain open. These include safety and emissions, congestion and many more. Further work will look at waiting times at stops, to better quantify the benefits of the proposed model. An interesting research problem could be studying the informal routes that the taxi drivers use for their day to day business in order to get passengers, or to avoid congestion on the main routes. More research needs to be done in the area of informal public transport sector to further improve its efficiency and safety.

VIII. Acknowledgement

The authors would like to acknowledge the following for their contribution during this research.

Microsoft Windows Azure for availing us with the computing resources to run some of the heavy memory consuming model tests.

Mr. Matsiko Pirez, the Technical Manager of Uganda Christian University for proving us with hosting service for our model server.

Mix Telematics for the data collection and hosting service.

IX. References

[1] United Nations economic commission for Africa. (2009, October, 30) *Africa Review Report on Transport.* (Sixth Session of the Committee on Food Security and Sustainable Development) Available: http://www.uneca.org/publications/africa-review-report-transport [June 09, 3014]

[2] M.J. Booysen. "Informal public transport in Sub-Saharan Africa as a vessel for Novel Intelligent Transport Systems" in *Proc. 16th International IEEE Annual Conference on Intelligent Transport Systems(ITSC 2013)*, 2013, pp 767-772

[3] A. Kumar and F. Barrett (2008, Jan) *Stuck in Traffic: Urban Transport in Africa* (Report)

[4] D. Solomatine, L.M. See and R.J. Abrahart " Data-Driven Modelling: Concepts, Approaches and Experiences" in Practical Hydro informatics , Springer Berlin Heidelberg, 2008, pp 17-30 , ISBN 978-3-540-79880-4

[5] Hang Zhou, Peter Hatherly "An Adaptive Data Driven Model for Characterizing Rock Properties from Drilling Data" in *proc Robotics and Automation (ICRA), 2011 IEEE International Conference* , 13 May 2011, pp 1909 - 1915

[6] B. Dalla Chiara, "Role of automatic vehicle location systems and localization accuracy in freight transport: an analytical estimation of gained operational time" . *IET Intelligent Transport Systems* , 2010, Vol. 4, Iss. 4, pp. 365–374 doi: 10.1049/iet-its.2009.0138

[7] L. Quadrifoglio, M Maged. "A simulation study of demand responsive transit system design" *ScienceDirect Transportation Research Part A: Policy and Practice* Volume 42, Issue 4, May 2008, Pages 718–737

[8] K. Khan, D. Renman & et..al "DBSCAN: Past, Present and Future" in *2014 Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT)*, 19 Feb. 2014, pp 232 – 238, 978-1-4799-2258-1

[9] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.