The Unbanked and Poverty:
Predicting area-level socio-economic vulnerability from M-Money transactions

Gregor Engelmann, Gavin Smith, James Goulding
N/LAB, Nottingham University Business School
University of Nottingham
Nottingham, England
psxge@nottingham.ac.uk

Abstract—Emerging economies around the world are often characterized by governments and institutions struggling to keep key demographic data streams up to date. A demographic of interest particularly linked to social vulnerability is that of poverty and socio-economic status. The combination of mass call detail records (CDR) data with machine learning has recently been proposed as a way to obtain this data without the expense required by traditional census and household survey methods. Based on a sample of 330k mobile phone subscribers resident in Dar es Salaam, Tanzania (7.6m M-Money records, 450.2m call and SMS event logs) this paper demonstrates the improvements that can be made via an alternate data stream: M-Money transaction records. An alternative to traditional banking services, particularly utilized by citizens unable to obtain a bank account, M-Money transactions provide a currently unexplored but potentially more powerful data set held by the same telecommunication companies.

Comparing directly to CDR as used in prior work the results show that M-Money provides an increase in socio-demographic classification accuracy (average F1 score) from 65.9% (0.63) to 71.3% (0.7) at much finer-grained spatial regions than previously examined. Notably, the combined use of M-Money and CDR data only increases prediction accuracy (average F1 score) from 71.3% (0.7) to 72.3% (0.71), providing evidence that M-Money is informationally subsuming CDR data. The reasons for this and the importance/contributions of individual features are subsequently investigated.

Keywords-M-Money; M-Pesa; Poverty prediction; CDR;

I. INTRODUCTION

More than 330 million people in Africa are still living in extreme poverty. This generates increased risk not only of disease and malnutrition, but to humanitarian abuses such as modern slavery and female genital mutilation (FGM). A lack of good-quality, fine-grained data for assessment of such vulnerability lies at the heart of dealing with this problem; a relentless challenge for intergovernmental organizations trying to design effective, targeted interventions to tackle the issue [1, 2]. Traditionally, the most reliable way to estimate poverty and socio-economic status has been through censuses or household surveys. Yet this pathway has proven impractically expensive and time-consuming. Household surveys only provide data for extremely small sub-samples of the population; censuses are coarse and rapidly out-of-date, often conducted but once a decade.

The situation is worsened by the fact that over 2 billion people lack access to traditional 'brick and mortar' banking infrastructure in emerging economies. This lack of access forces them to rely on typically unsafe, inconvenient and costly informal financial services or cash. A solution, however, exists in the proliferation of cheap mobile phones. The financial service offerings now being provided by telecommunication providers is allowing a rapidly growing proportion of those people access to "financial services such as payments, transfers, insurance, savings and credit" [3, 4].

Mobile phones were first introduced in Tanzania in the mid-1990s with a single network offering by TiGo. Similar to many other emerging economies, mobile phone adoption in Tanzania grew rapidly and is now very high with over 73% of adults in Tanzania owning at least one feature phone or smart phone [5]. Thus, recent work in the literature has sought to use Call Detail Records (CDR) as a Big Data solution for producing socio-demographic insights, and acting as a viable, remote proxy to fine-grained surveys.

Another data stream promises to achieve significant improvements on prediction models underpinned by Call Record Data - anonymized Mobile Money transaction logs themselves. The uptake in M-Money services has been extremely high in regions such as East Africa, with a growth of over 5.5 million users in the first four years since the introduction of M-Pesa by Vodacom in 2008 and TiGo Pesa in 2010. By 2013 M-Money agents and outlets significantly outnumbered traditional ATM's in the region with 17,000 compared to only 1,117 ATM's in Tanzania alone [6].

This reflects a wider trend within emerging markets, where at least nine other countries have recently been found to have more M-Money accounts than bank accounts [3]. By now, over 50% of Tanzanian telecommunication users have registered for M-Money with over 35% of households having at least one M-Money user compared to only 2% having active bank accounts [6, 7, 8]. These developments highlight the vast potential for such data streams to provide alternative methods for generating demographic data in emerging economies, to identify vulnerability, and to support policy initiatives in environments undergoing rapid structural changes.

In this paper, we demonstrate the superiority of features extracted from M-Money versus CDR data in predicting socio-demographic status for Dar es Salaam, Tanzania. Our analysis leverages a range of regularity, diversity, activity and spatial features derived from 450.2m call and SMS event logs and 7.6m M-Money records to generate insight into area-level socio-economic status. Results show that performance of models derived from M-Money data far exceeds that of state-of-the-art CDR-based models.

II. BACKGROUND

The majority of previous work on estimating poverty using non-traditional survey and census data has been conducted using satellite imagery. Such imagery allows for the collection of a range of derived data such as Night Time Lights (NTL), vegetation cover [9], proximity to important services (e.g. hospitals and schools) and infrastructure (e.g. density, proximity to paved roads). Advantages of satellite imagery include global coverage, a high revisiting capability and relative ease of access. NTL, in particular, has received a lot of attention, showing a good correlation with a country's GDP [10, 11]. Recent work in Bangladesh, however, suggests that as urban areas are nearing saturation levels of NTL and consequently the value of NTL as an indicator is beginning to decrease [12]. Moreover, analysis of earth observation imagery is both expensive and static, contrasting sharply with dynamic and digitally logged behavioural data streams such as Call Record Data.

A. Poverty mapping using CDR data

A number of previous studies have consequently utilized aggregate CDR data for poverty prediction. One of the first was [13], which combined a Support Vector Machine (SVM) predictor and CDR data for 500k users to analyze socio-economic levels within an urban area in a Latin American city. Follow on studies have analyzed CDR data in conjunction with data from phone surveys using linear regression [14] or in conjunction with environmental data using Gaussian Process Regression [16]. The spatial resolution varied from sub-prefecture and lower administration level in Cote d'Ivoire [12], Democratic and Health Survey (DHS) clusters[1] in Rwanda [14], and regions and communes in Senegal [15, 16]. An annotated bibliography is provided in table I, summarizing details and results of the key research papers on poverty mapping using CDR data.

Two issues arise in the prior literature: (1) a gap in conducting poverty prediction in small geographic areas such as wards or Lower Super Output Areas (LSOA) in part due to a lack in socio-economic data at such a fine granularity; and (2) a reliance on communication rather than mobile financial service data. To date, no previous studies exist that use features extracted from M-Money transaction logs, a data

stream that one would expect to provide significant insight into a fine-grained socio-economic analysis.

B. M-Money

Existing research on M-Money has mostly been limited to interview and survey-based work that broadly falls under three areas. First, research has examined the impact of M-money introduction, exploring people's usage patterns and the benefits they accrued. It was found, that M-Money has security advantages over traditional 'hiding places' for financial resources [7] and a reduced time commitment, costs and paperwork for initiating transfers compared to traditional banking services [17]. There has also been work looking at M-Money adoption among dedicated user groups such as Mpogole *et al.* (2016) [17] who studied the adoption among students and attitude of businesses owners towards M-Money in Iringa, Tanzania.

Second, research has investigated the impact of M-Money services on rural and agricultural areas in terms of inclusive development [18, 19]. It was found, that the proliferation of M-Money services enhances financial inclusion for those living in poverty by providing them with access to faster transfers and more secure ways of transporting and storing money [20]. Transporting money in a more secure way is particularly important in remote areas, where households often rely on remittances from family members that have traditionally been transported by paying bus drivers or through other risky means [21].

Finally, work has been undertaken to examine the safety and security of M-Money services [22]. Compared to traditional hiding places, which are susceptible to theft or loss, M-Money allows significantly safer storage and savings options [6]. At the same time, however, insufficient trust and technical sophistication for both sender and receiver are non-price related barriers to increased service adoption [23, 8].

In summary, while the research has looked in detail at the impact and barriers to adoption of M-Money services, such analysis has been limited to ethnographic studies - rather than the analysis of raw M-Money data streams themselves.

III. DATA DESCRIPTION

To address this research gap, this study leverages three disparate data sets: mass CDR and M-Money datasets for the generation of regularity, diversity, activity and spatial features, and ground reference survey data for validation of classification performance. The following subsections will provide an overview of the three data sets used as part of the study.

A. CDR data

CDRs are generated as part of a telecommunication operator's day to day operations for billing and network optimization purposes. A CDR is logged every time a network event such as sending an SMS, making a phone

---

[1] a geographic unit designed to be comparable to a village

Table I: List of studies on poverty prediction using mobile phone data

Ref	Data Source	Model (# of features)	Sample size	Time period	Results, Pearsons R	Spatial resolution	Poverty measure	Region
[13]	CDR	Support vector machine (279)	500k	6 months	0.8	-	Socio-economic levels	Urban area in a Latin American city
[12]	CDR	Linear regression (OLS) (10)	5M and 928k	5 months and 6 weeks	-	255 sub-prefectures — 176 areas at the next administrative level down	IMF and census	Cote d'Ivoire and undisclosed region
[14]	CDR and phone survey	Linear regression (5088)	1.5M (CDR) and 856 (survey)	9 months	0.68	492 DHS Clusters	DHS composite wealth index	Rwanda
[15]	CDR	Linear regression (5)	9M and 150k	12 months	0.82	14 regions	MPI (OPHI)	Senegal
[16]	CDR and environmental data	Gaussian Process Regression	9.54M	12 months	0.91	552 communes	MPI	Senegal

call or using mobile data occurs. As such, they allow insight into both micro- and macro-patterns of human interaction, while allowing for the preservation of individual anonymity through spatial and temporal aggregation. The data used as part of this study covers a total of 450.2m call and SMS events for 330k mobile phone subscribers taking place across the Dar es Salaam region of Tanzania over a period of 122 days in the autumn of 2014 [2]. Each record used as part of this study includes a range of attributes including:

- **SIM identifier**: anonymized identifier for the SIM card
- **Timestamp**: when the network event occurred
- **Location ID**: identifies the base transceiver station (BTS), which was used to service the network event

Residential behaviour was deduced from these records, through calculation of the mode Base Transceiver Station (BTS) favoured by users between 10 pm and 6 am. By geo-referencing the locations of these mode BTS, an anonymized dataset corresponding to users located in the Dar Es Salaam region was extracted. Features pertaining to residents at each BTS were then aggregated, ensuring both strict privacy and for analysis purposes. By undertaking a Voronoi tessellation of each BTS, a set of irregularly shaped cells can be generated at an average size of $6.23km^2$. The subset of areas classed as predominantly Urban averaged at $2.77km^2$, making them significantly smaller than sub-prefectures or DHS clusters used in other studies discussed in Table I.

B. M-Money data

M-Money is an umbrella term for a range of services offered by network operators, which include "sending and receiving money, making savings deposits, bill payments, making non-cash payments and transferring money from ones mobile phone account to bank accounts and vice versa" [17, p.4]. Similar to the CDR data, the M-Money data has been made available to us for research by a large Tanzanian telecommunications provider. The data set contains a sample of the company's M-Money transaction records for regular mobile phone users (subscribers), businesses (agents) and the network operator itself. Specifically, we extracted 47.6m M-Money records of approximately 147k customers for the same 122 day period as covered by the CDR data in 2014.

Each record contained a number of attributes collected by the network operator as part of the day-to-day M-Money provision. Those attributes include:

- **SIM identifier**: anonymized identifier for the handset
- **Date**: timestamp of when the transaction occurred
- **Transaction amount**: total monetary amount for the transaction, including service charge
- **Event type**: the category of the good/ service purchased via the transaction
- **Subtype**: a categorization of the business which provided the good/service featured in the transaction
- **Error code**: indicator of transaction success/failure, denoting the cause if the latter
- **User type**: account type of the individual invoking the transaction (e.g. subscriber/agent)

We further augmented the extracted M-Money features with home locations extracted from the CDR data set described in the previous section. Only users based in Dar es Salaam, who also made use of CDR services were included in the study.

C. Ground Reference Data

Accurate and fine-grained ground reference data necessary for supervised machine learning is extremely hard to find in East Africa. Additionally, most ground reference data sets are dictated by zoning, which does not correspond to the shape of capture areas for data streams generated through mobile phone usage. In order to overcome this limitation, we made use of a custom ground reference survey conducted over a 2 month period between late 2015 and early 2016 as part of the EPSRC *Neodemographics* project[3] and in collaboration with the *Ramani Huria* program [24]. That study surveyed a range of social and demographic attributes for over 500 areas across Tanzania's largest city, Dar es Salaam. The labelled areas are contiguous, irregular and cover the city's whole extent, having been derived via a Voronoi tesselation of WGS84 locations of each BTS. Surveying was undertaken by local inhabitants, with areas being surveyed at least twice. The "overall socio-economic status" was chosen as the focus of our study, with each area being labelled as either 'very poor', 'poor', 'average', 'wealthy' or 'very wealthy'. This covariate served as the key target variable for our modelling and evaluation process.

---

[2]Due to both individual and commercial privacy, the anonymized data used as part of this study is not publicly available, and was provided to us through a partnership with a telecommunications provider in Tanzania with a high market penetration in the case study area of Dar Es Salaam

[3]"Neo-demographics: Opening Developing World Markets by Using Personal Data and Collaboration", EPSRC Reference: EP/L021080/1, 2015

IV. Experimental Method

In order to investigate the utility of M-Money versus CDR transaction logs for remote socio-economic analysis, a core prediction task was formulated. A core dataset of 517 regions from Dar es Salaam, Tanzania was identified, each area being labelled with a ground-truth socioeconomic status. Competing models were generated from either M-Money, CDR or combined data, and the performance of each model in predicting socioeconomic status was tested via a strict cross-validation methodology. Two testing scenarios were formed, one considering prediction for all areas from the survey (n=517), and the second considering only areas labelled as being predominantly residential (n=384). Initial exploratory analysis over a range of model classes indicated *Random Forest* as performing highly effectively for the given scenario. Given the ability or Random Forests to directly model multi-class classification problems, handle non-linear relationships, and their association with well understood and tractable variable importance measures [25], subsequent evaluation focused on this class of model.

To allow direct statistical comparison of the effectiveness of M-Money versus CDR features, 3 models were constructed for each scenario (producing six in total). For each scenario, Model 1 was trained using an input feature set drawn from CDR data, with features fastidiously engineered to correspond directly to those used in the most recent literature (see §IV-A for more details of this process). Model 2 was trained using an M-Money derived feature set. Many features in Model 2 echo those in Model 1 (to provide a fair comparison), despite being seeded by transactional logs denoting mobile financial service events rather than call or SMS events. A final predictor, Model 3, was trained using a combination of all features used in the previous models. The features derived in each model vary in capturing different regularity, diversity, activity and spatial patterns of behaviour across the population.

The performance of each model was then assessed using a regime of repeated random (stratified) sampling, to generate 30 training and test sets. In each case, stratified five-fold cross-validation was used in conjunction with the training set to determine the optimal hyper-parameters before the performance was measured on the held out test set. Finally, an extensive variable importance analysis was performed, via rigorous investigation of permutation importance and partial dependency plots. The three parts to our experimental evaluation are discussed in more detail below, with overall results detailed in §V.

A. Feature engineering

The prediction task underpinning our models was formulated as a ternary classification problem in order to overcome the uneven distribution of class memberships in the raw survey data. Output feature labels 'very poor' and 'poor' were merged, as were labels 'wealthy' and 'very wealthy',

resulting in a relatively balanced dataset (with 202/143 areas labelled as poor, 153/102 as average and 163/140 as wealthy classes for all/residential scenarios respectively).

To populate the input feature space of each model, 17 features were extracted from the CDR data (based on prior work), and 22 from the M-Money data. Each feature corresponds to the usage patterns of a specific BTS, with data being aggregated using the mode BTS locations of accounts between 10 pm and 6 am. Features were engineered in a strict privacy-preserving fashion, with data not only being rigorously anonymized but aggregated to a BTS level. Such aggregation provides the analysis with an additional layer of privacy provision (recent estimates indicate over 4.3m Dar residents, so the average number of people living in each of our analysis areas is >8000).

We omit traditional per-feature correlation analysis (nor were feature spaces subsequently compression/factorization prior to modelling), given that we are primarily concerned with the analysis of feature utility and all features were based on pre-collected and re-purposed data with negligible acquisition cost.

While features in each set are broadly comparable with each other, a number of features unique to each data set were generated. Examples for Model 1 are items such as *ratio of text and calls* and response delay in SMS conversations; for Model 2 items such as *average M-Money in/out*, *percentage defaulted* and *percentage balance checks*. A closer examination of feature breakdown is supplied below, with an overview of all features used provided in table II.

CDR derived features:
Features for Model 1 were derived from transactions generated by 330k users acting over the sampling period and classed as living in Dar Es Salaam. They include a range of features identified as important for poverty prediction as identified by Pokhriyal *et al.* [16]. The features can be broadly classified as falling under five different domains:

- basic activity: the total number of SMS, the number of active days and the ratio between call and text interactions.
- regularity: the mean and standard deviation of time between call interactions.
- diversity: the mean balance of contacts for text interactions; the mean interactions per contact for call interactions; the percentage Pareto interactions for call interactions, and the entropy of contacts for call interactions
- activity: the mean and standard deviation in response delay for text interactions; the percentage of initiated interactions for call, and the percentage of initiated interactions for call and text interactions
- spatial: the frequently used antennas; the total number of used antennas; and the entropy of antennas

Table II: List of features generated from CDR and M-Money data for area-level socioeconomic prediction

Feature (total no. of features)	Data source	Description
Basic Use		
Number of interactions (2)	CDR, M-Money	The total number of incoming and outgoing SMS events and M-Money transactions for a user.
Number of users (2)	CDR, M-Money	The total number of users within the area
Average transaction size (1)	M-Money	Size of an average M-Money transaction across both incoming or outgoing transactions.
Average M-Money in/out (2)	M-Money	Average monthly inflow and outflow over the study period.
Total M-Money in/out (2)	M-Money	Total inflow and outflow over the study period.
Active Days (2)	CDR, M-Money	The number of days during which the user was active.
Ratio of text and call interactions (1)	CDR	The ratio of text and call interactions.
Spending uptake (1)	M-Money	The total spend in an area divided by the M-Money uptake (number of M-Money users divided by the number of CDR users in an area).
Ratio of incoming and outgoing transactions (1)	M-Money	The ratio of incoming and outgoing M-Money transactions.
Regularity		
Inter event time (4)	CDR, M-Money	The inter event time between two records of the user. This feature is calculated as mean and SD for M-Money, mean for calls and SD for calls and SMS.
Monthly events (1)	M-Money	Average number of transactions per month.
Diversity		
Balance of contacts (2)	CDR, M-Money	The balance of interactions per contact. This feature is calculated-each for text and M-Money. For every contact, the balance is the number of outgoing interactions divided by the total number of interactions (in + out).
Interactions per contact (2)	CDR, M-Money	The number of interactions a user had with each of his or her contacts via call or M-Money
Percentage Pareto interactions (2)	CDR, M-Money	The percentage of user's contacts that account for 80 of his or her call or M-Money interactions
Entropy of contacts (2)	CDR, M-Money	The entropy of the user's contacts for calls or M-Money
Activity		
Response delay (2)	CDR	The response delay of the user within a conversation (in seconds). Calculated for text (SD and mean of the response delay).
Percentage initiated (3)	CDR, M-Money	The percentage of network events initiated by the user for calls, call and SMS, or M-Money.
Percentage defaulted (1)	M-Money	The percentage of transactions that failed due to insufficient account funds.
Percentage balance checks (1)	M-Money	The percentage of transactions representing balance checks.
Spatial		
Number of BTS (3)	CDR, M-Money	The number of unique cells or BTS visited.
Frequent BTS (1)	CDR	The number of BTS that accounts for 80 of locations where the user was.
Entropy of BTS (1)	CDR	The entropy of visited BTS.

M-Money derived features:
Features for Model 2 were derived from transactions covering 147k users with both incoming and outgoing M-Money transactions occurring in every month of the sampling period, and who were classed as likely living in Dar Es Salaam. Average M-Money in/out represents the amount of money, which was received or spent during an average month. Only transactions conducted by 'subscribers' were taken into account to calculate those features. Users with an average of more than 40 incoming events or 100 outgoing events per month were excluded from the analysis, as these most likely represented unlicensed businesses or informal street traders operating as regular subscribers. As error codes were changing over the course of the year, only transactions without error codes or those indicating success were included. Similarly, transaction amounts equal to or under 50TZS were excluded from the analysis, as they mostly referred to balance inquiries or pin changes, which involved minuscule service charges, that introduced significant noise into the M-Money feature calculation.

Additionally, we captured the percentage of defaulted transactions, and percentage of balance checks. Defaulted transactions were identified as having failed due to error codes indicating an insufficient account balance to complete out a particular transaction. Balance checks are identified as transactions with an event type for 'Balance inquiries'.

B. Evaluation setup

To evaluate the comparative utility of the CDR vs M-Money features in real-world applications we consider two predictive tasks (1) predicting the economic status of *only residential* regions populations and (2) predicting the economic status of *all* regions populations. The first represents the more focused task of interest and is based on the collected ground-reference survey (see §III-C), while the second represents the task if there is no knowledge as to the residential/non-residential status of regions. In each case, a combined model with all features (CDR + M-Money) and two models per data source (one only containing CDR features, one only containing M-Money features) was evaluated. Output labels were generated based on the ground-reference data set discussed in §III-C, with each region labelled as either 'poor', 'average' or 'wealthy'.

For each model, the data was split via stratified random sampling into a training (66%) and test set (33%) and the parameters for the Random Forest (number of trees, max-

imum depth and the minimum samples per split) selected via a grid search underpinned by stratified five-fold cross-validation using only the training set and the model finally trained with the best parameters. The performance of each model was then tested on the held out test set. This was repeated 30 times arriving at 30 performance scores for each of the six models.

C. Evaluation criteria

The performance of each model was measured via the precision, recall and the F1 score (harmonic mean of the precision and recall) for each of the 30 runs per model. The average results over all runs are shown in Table III, along with classification accuracy. Box-plots showing the per-class distribution of the overall F1 scores per model are shown in Fig 1a for all areas and Fig 1b for residential areas only.

D. Analysis of Variable Importance

Of equal, if not greater importance than overall model accuracy, is the ability to break apart our models to investigate the work being done by individual CDR and M-Money features. To achieve this, the final step in our investigation was a variable importance analysis, consisting of three stages. First, and due to a large number of features proposed in the literature (and included in this work), our candidate set was filtered to exclude those features shown to have minimal impact on model performance. Such features were identified using the *Boruta algorithm* [26, 27], which only considers variables as important if they provide a statistically significant increase in prediction strength compared to a permuted (i.e. randomized) version of the feature. We set the level required to achieve statistical significance low ($p = 0.1$) to minimize the probability of inadvertently discarding relevant features.

Second, surviving features were used to fit optimized Random Forest models for the whole data-set, selecting optimized parameters via a grid search. The out-of-bag[4] (OOB) F1 score was then checked to ensure the reduction in features did not cause a significant decrease in overall generalized performance. This then allowed a full permutation importance analysis to be undertaken. Permutation importance was used due to both its interpretability (illustrating the mean decrease in performance of omitting each feature) as well as its ability to attribute variance in the case of non-linear interactions. With important features thus identified, their behaviour within the model and across sub-populations of the data was further examined using partial dependence plots in §V-C [29, 30].

---

[4]Out-of-bag refers to the use of only the trees in the forest for which the sample being predicted was not part of during training. In this way, the resulting performance measure can be considered to represent the generalized error [28].

V. Results and Discussion

A. Model Performance Results

Table III reports all results for our experiments, detailing the F1, precision and recall scores averaged over each of the 30 experimental runs, accompanied by an overarching classification accuracy score. As can be seen, in all scenarios those models containing M-Money features strongly outperform those that use CDR features alone. The final column, indicating per-class F1 scores for each model, illustrates that this improvement is not due to one class being favoured, but occurs across the board - and in particular improves prediction of average, middle-income areas.

Of note is the improvement in F1 scores of Model 1 against Model 2, with an increase in accuracy of ~9.7% for all areas, and ~9.2% for residential areas. In both cases, while we can observe an increase in the performance of models containing M-Money features, there are only marginal differences in using combined data sets. These results indicate that the M-Money features appear to be a better indicator of underlying socio-economic trends compared to baseline CDR features. Further, it provides evidence that M-Money features are subsuming the information contained in CDR features when it comes to poverty/socio-demographic classification, and CDR is likely capturing only a subset of the variance covered by the M-Money feature set.

Models trained on residential-only data clearly function more effectively than those also attempting to predict the level of poverty in non-residential areas (improving by 1.4%, 2.1% and 1.1% for Models 1-3 respectively), highlighting the difficulty in categorizing non-residential areas with a socio-demographic label.

The scope of improvements made by models leveraging M-Money features is illustrated in Figures 1a and 1b, which show box plots of per-class F1 scores for all 3 feature sets in both scenarios. We can observe in Figure 1a that F1 scores for middle-income areas (0.51) are on average ~39% lower than scores for poor (0.76) and ~42% for wealthy (0.78) areas. A similar, albeit even more pronounced trend, can be observed in the residential scenario with a low F1 score for average (0.46) and a difference in ~50% to poor (0.77) and ~54% to wealthy (0.8) areas. Predicting average areas is tricky, with most model features apparently only supporting a binary poor/not-poor decision boundary.

Nevertheless, it is for middle-income areas that M-Money features appear to make the most gains over CDR, with a large gap between Model 1 (0.43) and Model 2's performance (0.54) across all areas. It is clear from the box plot that the variance of CDR features in middle-income areas is extremely wide, rendering them ineffective in delineating 'average' areas from any other.

B. Variable Importance

Given the advantage that experiments indicate M-Money features have over CDR, consideration turns to why this

(a) Per-class F1 scores across all BTS areas in the dataset

(b) Per-class F1 scores across all Residential areas in the dataset

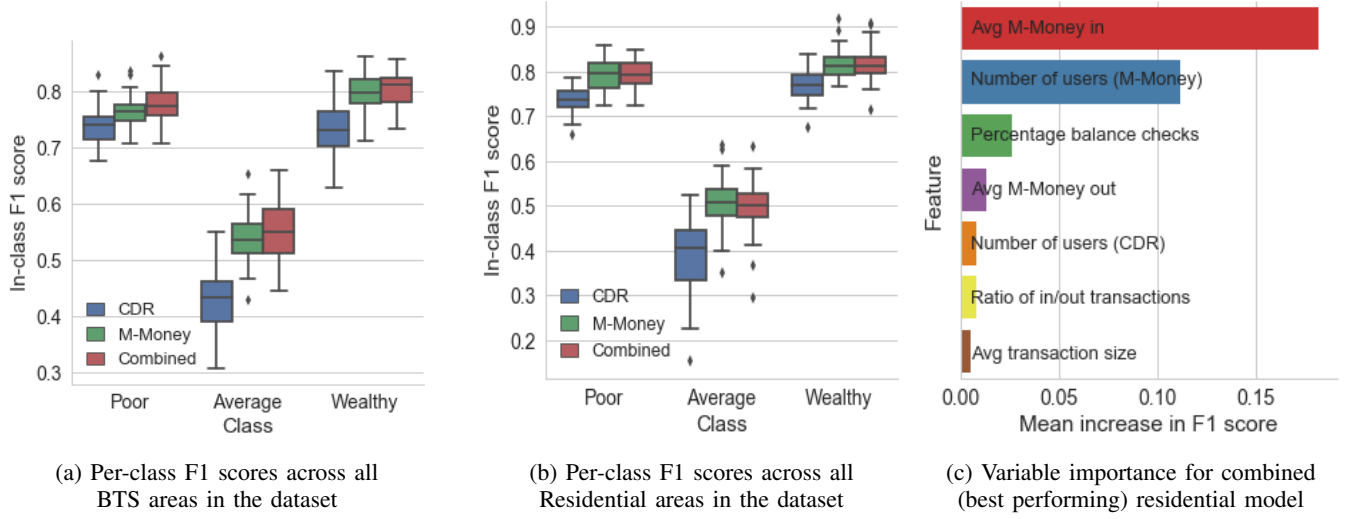(c) Variable importance for combined (best performing) residential model

Figure 1: Subfigures (a) and (b) illustrate the significant improvements made across all 3 classes (*poor*, *average*, *wealthy*) using M-Money rather than CDR features. Variable importances in (c) are for a combined CDR/M-Money model.

Table III: Accuracy Results for random forest prediction using different feature sets over 30 randomly seeded experimental runs

Model Feature Set	Model Performance	Classification Accuracy	Per-class F1 score
CDR features	f1 score: 0.63 precision: 0.64 recall: 0.64	65.9%	Poor: 0.74 Average: 0.43 Wealthy: 0.74
M-Money features	f1 score: 0.7 precision: 0.71 recall: 0.7	71.3%	Poor: 0.77 Average: 0.54 Wealthy: 0.8
Combined (CDR + M-Money)	f1 score: 0.71 precision: 0.72 recall: 0.71	72.3%	Poor: 0.78 Average: 0.55 Wealthy: 0.8
CDR features residential	f1 score: 0.63 precision: 0.64 recall: 0.64	67.2%	Poor: 0.74 Average: 0.38 Wealthy: 0.77
M-Money features residential	f1 score: 0.71 precision: 0.71 recall: 0.71	73.4%	Poor: 0.79 Average: 0.51 Wealthy: 0.82
Combined residential	f1 score: 0.7 precision: 0.71 recall: 0.71	73.4%	Poor: 0.79 Average: 0.49 Wealthy: 0.82

is occurring. To this end, permutation importance was investigated as per the methodology described in §IV-D. For each model 1 feature within the test set (OOB samples) at a time is permuted (simulating holding out that feature without requiring a full model training) and the performance (F1 score) compared to the performance obtained with the full model. Note that the use of OOB samples provides the variable importance of the generalized predictor, rather than fitted (potentially overfit) model. Results for the combined

feature set model are shown in Figure 1c, with the x-axis being interpreted as the mean increase in F1 score that the inclusion of the variable provides (assuming all other features are kept in the model). Six out of the seven most predictive features are generated via M-Money data (in fact the only useful CDR feature relates to *number of users*, which is likely to reflect the population density in an area and can be drawn from other sources). The top 4 M-Money features overall are: *average M-Money in*; *active M-Money users*; *percentage balance checks*; and *average M-Money spend*. The average M-Money received by residents across an area clearly explains the most variance, contributing a 0.18 increase in F1 score. This is followed by the number of M-Money users in second at 0.11; the percentage balance checks at 0.03; and the average M-Money out at 0.01.

C. Understanding M-Money Feature Effects

To break down the feature effects further we use *Partial Dependence Plots* (PDPs) to visualize the increase/decrease of the probability of predicting a given output class (i.e. wealthy) when a factor is varied while all others are kept fixed. PDPs are highly effective in showing us the model's sensitivity to the feature in question and how its predictions will respond as the variable's value changes. In each diagram, the heavily weighted line shows the mean change in the probability of a data-point being assigned to a particular class, as the variable increases. Considered in conjunction with the permutation importance scores, the PDPs highlight the nature of the relationship between the M-Money variables and socio-economic class.

As one might expect, the model clearly identifies that as the amount of M-Money incoming to residents of a BTS increases the likelihood of that area being affluent also
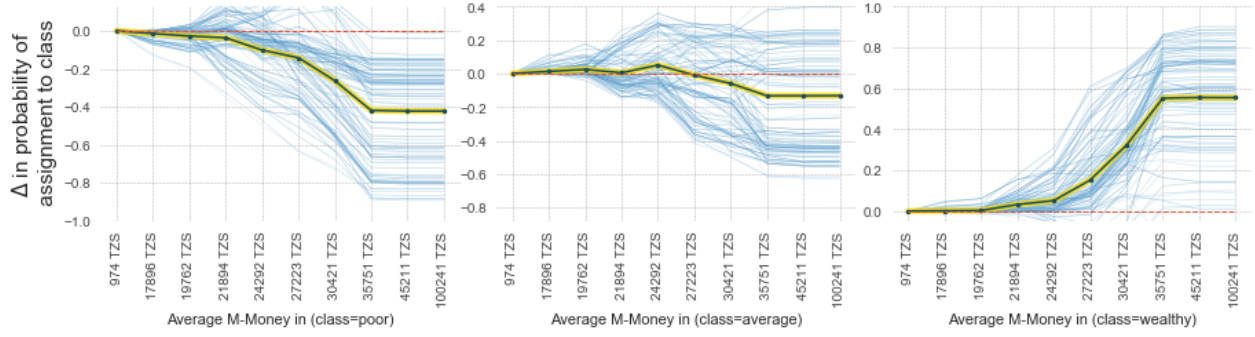
Figure 2: Partial Dependence Graphs for Average M-Money received by residents in each area of Dar es Salaam

increases (as denoted by the large areas under the curve for Figure 3a and c). However, the feature provides minimal information to the model for categorization for 'average' areas. There is, in fact, a relatively wide variance of *avg money in* in average areas (which can be both middle-income areas or zones with combined informal/residential housing). Thus, we are left with a feature which provides binary classification - figure 1a and 1c being mirror images of each other. There is a plateau to the informativeness of this effect, however, and as a predictor, its partitioning effectiveness peaks at ~35,000 TZS (which thus might well define a heuristic boundary between poor/wealthy users).

The number of residents who use mobile financial services in an area is negatively correlated with affluence, as illustrated in Figure 3c, where higher uptake of M-Money increases the probability of a wealthy classification. This may initially seem counter-intuitive, but it is the unbanked that have the highest propensity to need an alternative to traditional financial services. While the poor remain disengaged from credit card usage, mobile phone usage is ubiquitous, even in slum areas, resulting in this effect.

The most important feature to the model in classifying middle-income towers is average M-Money spend. It is this feature, which is likely improving the most over CDR data in the assessment of such areas. Figure 4 shows that while the model broadly associates an average M-Money spend of ~25,000 TZS as the cut off line for poorer areas, it is only at ~29,000 TZS where the likelihood of assignment as a rich area begins to jump. In between, the likelihood of a middle-income area being assigned increases. However, the feature is not decisive, and the true situation remains blurry to the model with two broad trends visible in figure 4b. Half of the blue lines form a hump in the middle of the graph (reflecting a strict middle-income sub-population perhaps) dropping once spend gets over a certain threshold, while the other half remain high (and thus will not be distinguished from residents of a wealthy area). Such fuzziness is likely due to the different types of average areas that occur in reality - true middle-income areas, and those with distinct mixtures of wealthy and less affluent communities.

A further useful feature identified by the model is the average of balance checks users make to their mobile money accounts. As balance checks increase there is more likelihood of an area correctly being assigned as poor (although this is not a monotonic relationship). This is likely to be reflecting the fact, those living closer to the breadline need to assess their exact financial situation far more regularly than those who are more affluent.

VI. STUDY LIMITATIONS

1) User selection: The study includes only individuals using basic network services (M-money, SMS, calls). Geographic indicators were not present in M-Money logs (although this is a symptom of our sample, rather than the raw data). Furthermore, user types (see Figure 1a) were provided by the telco company, and accuracy was assumed - on the ground analysis, however, has indicated a large informal economy of street traders operating under the auspice of being regular subscribers across the region, and this ought be remembered when considering feature importance.

There are also potential issues around selection bias arising from mobile phone ownership among different groups of society. In urban spaces in Tanzania, however, mobile phone penetration has been reported to be close to 92% resulting in a negligible uptake bias. Furthermore, the telco providing the study's datasets has a 28% per cent market share, with over 70% of overall customers reporting that they had not used another network [7].

2) Validation data: Supervised machine learning techniques require accurate training and validation samples. As part of this study, we made use of a custom ground reference survey collected by volunteers within Dar Es Salaam over a 2-month long period. While the sampled areas conformed to the same BTS derived Voronoi tessellation, assignment of these areas into clear socio-demographic categories is often difficult (as shown in a sample satellite image in figure 5). There remains the potential for errors in surveyed data, and this issue is exacerbated in 1. the rural and peri-urban outskirts of Dar Es Salaam, where large coverage areas of individual BTS are encompassing a mixture of affluence
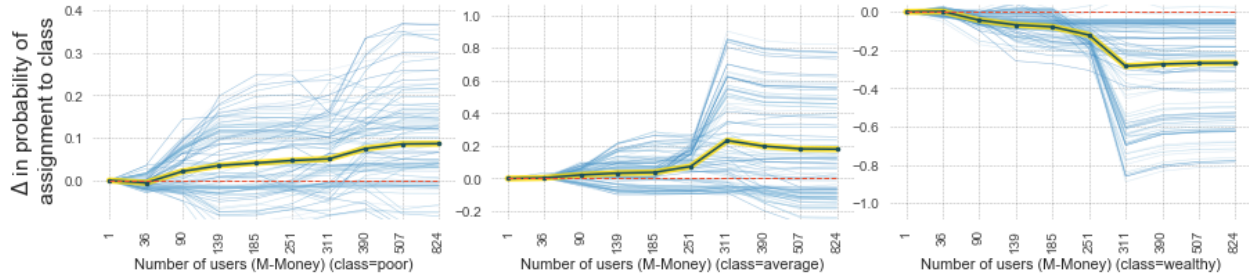
Figure 3: Partial Dependence Graphs for the number of M-Money users in each area of Dar es Salaam
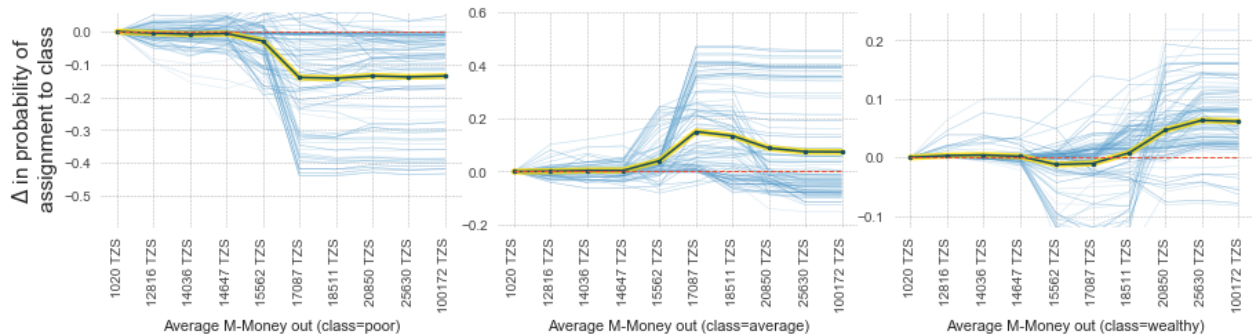

Figure 4: Partial Dependence Graphs for Average M-Money spend by residents in each area of Dar es Salaam

Figure 5: Sample area within Dar. A slum occurs in the centre with affluent housing left and right of the corridor

levels, and 2. the dense makeup of the city, where slums and formal residential areas can sometimes be found in extremely close proximity.

3) Data accessibility: Continuing, rigorous care must be taken in examination of CDR data and Mobile Money data. While such datasets are generated and curated by commercial telcos, they contain extensive personal data. While initial guidelines for best-practice engagement with third-party organizations to analyze CDR data have been developed by the GSMA [31] (in the wake of the recent Ebola crisis), overarching policies are yet to be established in regions where research can be of the most benefit. This remains key to ensure the strict preservation of both indi-

vidual and commercial privacy while allowing the potential of leveraging 'big data' for social good to be fully realized.

VII. CONCLUSION

Using ground reference, CDR and M-Money data, we identified and quantified features with the aim of accurately classifying which small-scale areas across the Tanzanian port city of Dar Es Salaam fall within different socioeconomic categories. Our approach compared baseline metrics extracted from CDR data in line with previous approaches on harnessing CDR data for area-level poverty assessment discussed in table I with features extracted from M-Money data, combinations of those feature sets, and feature sets incorporating a differentiation in underlying land-use.

This study demonstrates that mass M-Money datasets can provide sizable improvements in socio-demographic classification accuracy over recent CDR approaches. However, limitations still remain, largely due to a need for more high-quality ground reference data for training. At the same time, we believe this work will provide useful information to empower policymakers and local municipalities to identify areas requiring interventions and revitalization programs, as according to the UNFPA, "any indicative estimates would provide in certain situations where none are currently available; even if they carried with them a significant level of uncertainty such estimates would still represent a large improvement in many cases" [12, p. 519].

REFERENCES

[1] S. Devarajan, "Africa's statistical tragedy," *The Review of Income and Wealth*, vol. 59, 2013.

[2] M. Jerven, "Comparability of gdp estimates in sub-saharan africa: The effect of revisions in sources and methods since structural adjustment," *The Review of Income and Wealth*, vol. 59, 2013.

[3] GSMA, "State of the industry: Mobile financial services for the unbanked," GSMA, Tech. Rep., 2014.

[4] V. Mishra and S. Bisht, "Mobile banking in a developing economy: a customer-centric model for policy formulation," *Telecommunications Policy*, vol. 37, 2013.

[5] PewResearchCenter, "Cell phones in africa: Communication lifeline," 2015.

[6] NetHope and MEDA, "Tanzania mobile money assessment and case study: Examining cash payment streams and their electronic alternatives among usaid implementing partners," Tech. Rep., 2013.

[7] InterMedia, "Mobile money in tanzania: Use, barriers, and opportunities. the financial inclusiontracker survey project," Tech. Rep., 2013.

[8] N. Economides and P. Jeziorski, "Mobile money in tanzania," *Marketing Science*, vol. 36, 2017.

[9] P. O. Okwi, G. Ndeng'e, P. Kristjanson, M. Arunga, A. Notenbaert, A. Omolo, N. Henninger, T. Benson, P. Kariuki, and J. Owuor, "Spatial determinants of poverty in rural kenya," *PNAS*, vol. 104, 2007.

[10] S. Ebener, C. Murray, A. Tandon, and C. C. Elvidge, "From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery," *International Journal of Health Geographics*, vol. 4, 2005.

[11] C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, and E. R. Davis, "Mapping city lights with nighttime data from the dmsp operational linescan system," *Photogrammetric Engineering and Remote Sensing*, vol. 63, 1997.

[12] C. Smith-Clarke, A. Mashhadi, and L. Capra, "Poverty on the cheap," in *CHI 2014, One of a CHInd*, Toronto, CA, 2014.

[13] V. Soto, V. Frias-Martinez, J. Virseda, and E. Frias-Martinez, "Prediction of socioeconomic levels using cell phone record," in *19th Int. Conference on User Modeling, Adaption and Personalization*, Gerona, Spain, 2011.

[14] J. Blumenstock, G. Cadamuro, and R. On, "Predicting poverty and wealth from mobile phone metadata," *Science*, vol. 350, 2015.

[15] N. Pokhriyal and W. Dong, "Virtual network and poverty analysis in senegal," in *NetMob Book of Abstracts*, MIT Media Labs, Boston, 2015.

[16] N. Pokhriyal and D. C. Jacques, "Combining disparate data sources for improved poverty prediction and mapping," *PNAS*, vol. 114, 2017.

[17] H. Mpogole, Y. Tweve, N. Mwakatobe, and D. S. Serijo Mlasu, "Towards non-cash payments in tanzania: the role of mobile phone money services," in *IST-Africa 2016 Conference*, 2016.

[18] C. Qiang, S. Kuek, A. Dymond, and S. Esselaar, "Mobile applications for agricultural and rural development," Tech. Rep., 2011.

[19] S. Asongu and N. M. Odhiambo, "Mobile banking usage, quality of growth, inequality and poverty in developing countries," Tech. Rep., 2017.

[20] NetHope, "Making the journey from cash to electronic payments," Tech. Rep., 2013.

[21] W. Jack, A. Ray, and T. Suri, "Transaction networks: Evidence from mobile money in kenya," *American Economic Review*, vol. 103, 2013.

[22] R. Gwahula, "Risks and barriers associated with mobile money transactions in tanzania," *Business Management and Strategy*, vol. 7, 2016.

[23] N. Mallat, "Exploring consumer adoption of mobile payments a qualitative study." *The Journal of Strategic Information Systems*, vol. 16, 2007.

[24] Dar Ramani Huria, http://ramanihuria.org/.

[25] W. R. Rudnicki, M. Wrzesień, and W. Paja, *All Relevant Feature Selection Methods and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015.

[26] M. B. Kursa and W. R. Rudnicki, "Feature selection with the boruta package," *Journal of Statistical Software*, vol. 36, 2010.

[27] H. Stoppiglia, G. Dreyfus, R. Dubois, and Y. Oussar, "Ranking a random feature for variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1399–1414, 2003.

[28] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013.

[29] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[30] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *Journal of Computational and Graphical Statistics*, vol. 24, no. 1, pp. 44–65, 2015.

[31] GSMA, "Gsma guidelines on the protection of privacy in the use of mobile phone data for responding to the ebola outbreak," 2014.