# Predictors without Borders: Behavioral Modeling of Product Adoption in Three Developing Countries

Muhammad Raza Khan
University of Washington
Seattle, WA 98195
mraza@uw.edu

Joshua E Blumenstock*
University of Washington
Seattle, WA 98195
joshblum@uw.edu

## ABSTRACT

Billions of people around the world live without access to banks or other formal financial institutions. In the past several years, many mobile operators have launched "Mobile Money" platforms that deliver basic financial services over the mobile phone network. While many believe that these services can improve the lives of the poor, in many countries adoption of Mobile Money still remains anemic. In this paper, we develop a predictive model of Mobile Money adoption that uses billions of mobile phone communications records to understand the behavioral determinants of adoption. We describe a novel approach to feature engineering that uses a Deterministic Finite Automaton to construct thousands of behavioral metrics of phone use from a concise set of recursive rules. These features provide the foundation for a predictive model that is tested on mobile phone operators logs from Ghana, Pakistan, and Zambia, three very different developing-country contexts. The results highlight the key correlates of Mobile Money use in each country, as well as the potential for such methods to predict and drive adoption. More generally, our analysis provides insight into the extent to which homogenized supervised learning methods can generalize across geographic contexts. We find that without careful tuning, a model that performs very well in one country frequently does not generalize to another.

## CCS Concepts

**•Information systems → Data mining; •Computing methodologies → Supervised learning by classification; •Theory of computation →** *Automata extensions;* **•Applied computing →** *Economics; Sociology;*

## Keywords

Mobile Money; Feature engineering; Gradient boosting; Product adoption; Supervised learning

*Correspondence should be addressed to joshblum@uw.edu



*KDD '16 San Francisco, California USA*
© 2016 ACM. ISBN 123-4567-24-567/08/06...$15.00
DOI: 10.475/123_4

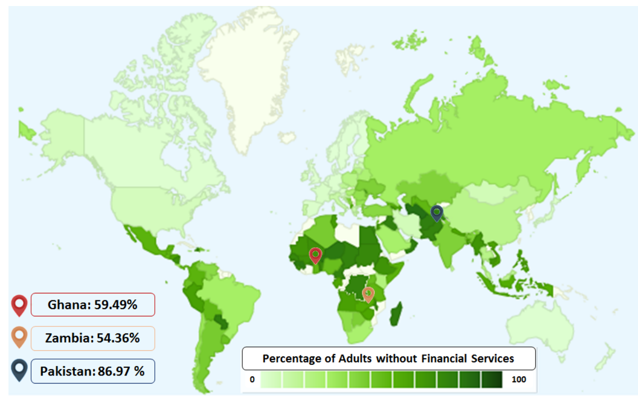


**Figure 1: Worldwide access to formal financial services, constructed using data from the Global Financial Inclusion Database [7]. Study locations are identified by pins.**

## 1. INTRODUCTION

The rapid penetration of mobile phones in developing countries is creating new opportunities to provide basic financial services to billions of individuals who have never before had access to banks or other formal financial institutions (Figure 1). In particular, over the last several years, mobile phone operators across the globe have launched "Mobile Money" platforms, which make it possible for mobile phone subscribers to conduct basic financial transactions from inexpensive feature phones. In several countries, these platforms have been wildly successful: two thirds of all Kenyan adults are active subscribers on the dominant Kenyan Mobile Money system [26]; in Bangladesh and Tanzania the corresponding usage rates are 40% and 50% [6, 9]. Globally, there are 255 live Mobile Money deployments in 89 countries, with an additional 102 planned deployments in the near future. With industry group GSMA estimating that 1 billion individuals currently own a phone but do not have a bank account [27], this presents massive potential to provide useful services to poor customers.

However, outside of the countries mentioned above and a few others, worldwide adoption of Mobile Money has been extremely anemic. The vast majority of deployments have struggled to promote sustained product adoption, and an industry report from 2014 estimates that 66% of registered customers were inactive [27]. An open and important question thus revolves around understanding what drives cus-

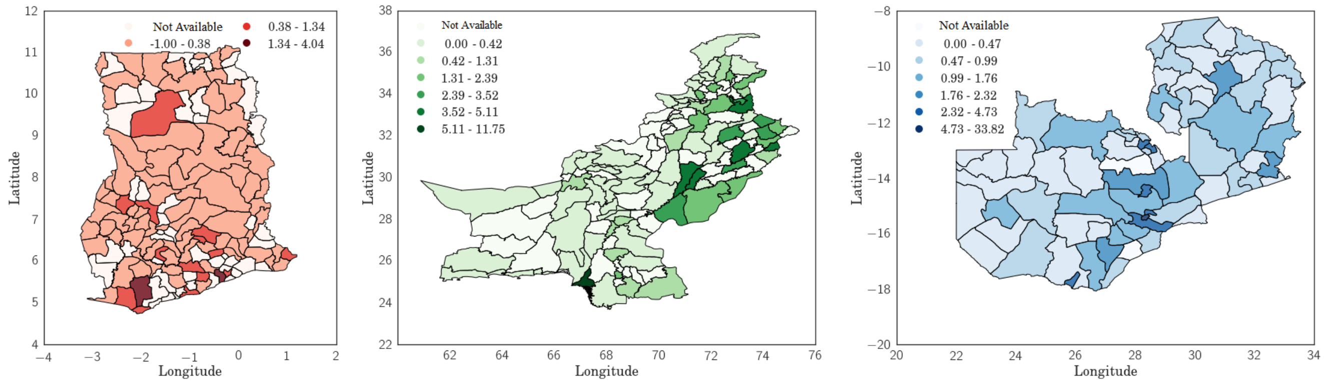


**Figure 2: Geographic distribution of registered Mobile Money users in Ghana, Pakistan, and Zambia. Cells are colored according to the fraction of Mobile Money users in each region**

tomers to adopt and use Mobile Money, and whether patterns observed in one country will generalize to another.

In this paper, we use spatio-temporal transactions data on mobile phone activity to model Mobile Money adoption in three developing countries. Our focus is on mining the Call Detail Records (CDR) collected by mobile phone operators, which contain detailed metadata on all events that transpire on the mobile phone network, including phone calls, text messages, and Mobile Money transactions. For thousands of unique individuals in each country, we can thus infer a wealth of information about the structure of their social networks, their daily movements about the country, patterns of communication, and several other behaviors that we discuss in greater detail below. We also know whether each subscribers eventually signs up for Mobile Money, and if so, whether he or she remains an active user on the system. In the three countries we study - Ghana, Pakistan, and Zambia - each Mobile Money platform is owned and operated by a separate, independent mobile phone operator, and the subscriber population in each country has very different social and economic characteristics.

There are three substantive and one methodological contributions of this study. Substantively, we (1) develop a richer understanding of what drives the adoption of Mobile Money, by mining several large databases of transactions data; (2) construct a supervised learner that can predict, to varying degrees of accuracy depending on the prediction task and country context, the likelihood than an individual subscriber will use Mobile Money; and (3) explore the possibility that transfer learning could be used to train models in one country or context and apply them in another. To our knowledge, this is the first study to train and evaluate models of product adoption in three very different contexts. Since Mobile Money adoption is notoriously idiosyncratic, we hope this "cross-cultural" comparison can provide insight into the generalizability of our results, and increase their broader relevance to the policy and business communities working in developing countries.

Methodologically, we develop a novel framework for extracting behavioral metrics from transaction logs, which produces interpretable features that can provide the input data into standard supervised learning algorithms. This framework extends previous efforts described in [2], which used a simplified approach to predict poverty and wealth from mobile phone data. The core of this approach is formalized as a Deterministic Finite Automaton, which provides a structured, recursive grammar that relies on relatively few degrees of freedom to generate a comprehensive and interpretable set of "dense" features from sparse log data. This approach is sufficiently generalizable that we hope it can be further extended to a much broader range of contexts where researchers and data scientists wish to extract interpretable knowledge from transaction log data.

## Related Work

Our work builds on several distinct strands in the academic literature. The first is concerned with understanding the determinants of mobile money adoption. This literature has historically been the domain of development researchers, and includes both macroeconomic and ethnographic work. The macro-scale work is concerned with the national and regulatory forces that can promote and hinder the spread of mobile money, such as interoperability regulations, barriers to customer registration, and the need for a robust network of mobile money agents [21, 8, 11]. The ethnographic work has focused primarily on qualitative studies of how mobile money can be integrated into the daily lives of the poor [24, 23, 12].[1]

A second strand of literature seeks to derive general insights from patterns revealed in mobile phone transactions logs. This encompasses a wide array of applications, including predicting the socioeconomic status [2], gender [13], and age [10] of individual mobile phone subscribers. These studies illustrate the rich signal latent in mobile operator data, which reflects social phenomena including the structure of social networks, patterns of mobility and migration, diurnal rhythms of daily activity, and expenditures on communication and airtime.

A third area of prior work, and the one most relevant to our study, contains several papers that use transactions data from mobile operators to study product adoption.[2] For in-

[1] A closely related body of work explores the welfare consequences of the spread of Mobile Money [1, 18, 3, 4], though relatively few studies provide rigorous evidence that mobile money has a positive impact on the lives of the poor.

[2] A much broader literature, which we do not review here, studies the role of social networks in the adoption of new technologies [29, 20].

| Country | Ghana | Pakistan | Zambia |
|---|---|---|---|
| *Panel A: National statistics (Source: World Bank)* | | | |
| Population | 25.90 Million | 185.00 Million | 15.72 Million |
| Percent with bank accounts | 40.51 | 3.02 | 45.64 |
| GDP per capita (PPP adjusted) | \$4081.70 | \$4811.4 | \$3904.00 |
| Mobile phone subscriptions (per 100 people) | 115 | 73 | 67 |
| Mobile phone operators | 6 | 6 | 3 |
| *Panel B: Mobile phone use (Source: Call Detail Records)* | | | |
| Calls per subscriber per day | 6.53 (6.99) | 7.76 (10.25) | 10.26 (102.86) |
| SMS per subscriber per day | 3.10 (100.36) | 38.71 (80.83) | 10.88 (262.81) |
| Number of unique contacts | 21.66 (24.91) | 46.93 (139.67) | 17.63 (328.63) |
| Number of unique Towers | 12.98 (16.07) | 24.15 (57.30) | 7.35 (17.56) |

*Notes*: Standard deviations reported in parenthesis.

**Table 1: Summary statistics by country: national indicators and sample CDR metrics**

stance, Khan et al. [19] use billing data to predict customer churn in an Afghan telecom, using a brute-force approach to feature generation. Sundsoy et al. [28] construct 350 features from call detail records and compare the performance of basic machine learning algorithms to that of a marketing department in predicting uptake of data plans. Finally, an industry report by CGAP compares the relative influence of different types of mobile phone metrics on the adoption of Mobile Money in Africa, using 180 metrics derived from call data [5]. Relative to these studies, our study moves this literature forward by (a) innovating in the method used to generate features, thereby providing a systematic and comprehensive approach to feature engineering; (b) leveraging data from three different contexts to calibrate the external validity and generalizability of our results; and (c) carefully articulating the experimental protocols and algorithms in a way that will enable other researchers to replicate and extend these methods.

## 2. DATA AND CONTEXT

For this study, we worked in collaboration with three mobile phone operators in Ghana, Pakistan, and Zambia. All three countries rank in the bottom third of the Human Development Index, a metric developed by the United Nations to capture a broad range of welfare outcomes such as income, education, inequality, and life expectancy. As can be seen in Figure 1, penetration of financial services is very low in each country. Geographic patterns of Mobile Money adoption also vary greatly within each country, as shown in Figure 2. Additional information on each country is provided in Table 1. Of course, these country-level statistics mask enormous diversity between and within nations in social and demographic characteristics, religious and political attitudes, and general ways of living.

From each mobile phone operator, we obtained the anonymized Call Detail Records (CDR) and Mobile Money Transaction Records (MMTR) of every subscriber on the network. The CDR and MMTR contain basic metadata on every event that occurs on the mobile phone network, including phone calls, text messages, and any form of Mobile Money activity. CDR typically consists of tuples containing `{callerID, recipientID, date, time, duration, callerLocation}`, where the two ID's are anonymized phone numbers, the date and time indicate when the event transpired, the duration of the call is recorded in seconds, and the location

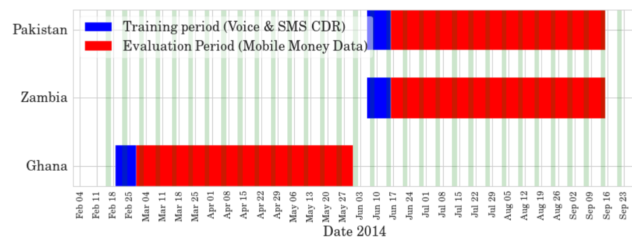


**Figure 3: Training and evaluation periods**

field indicates the cellular tower through which the call was routed, which can be used to pinpoint the approximate location of the individual at the time of the call.[3] For the mobile money platforms we study, the MMTR contain similar metadata for basic financial transactions, such as deposits, withdrawals, purchases, balance checks, and so forth.

In total, the original data contains billions of transactions conducted by tens of millions of unique individuals. Each dataset spans several months of activity, which we divide into a "training" period and an "evaluation" period. CDR from a 10-day training period was used to engineer features and fit a predictive model, where the target variables (based on Mobile Money activity) were measured in a subsequent 3-month evaluation period. The timing of these periods is depicted in Figure 3.[4]

Using data from the evaluation period, each subscriber in each sample was labelled as either a "Voice Only" User or a "Registered Mobile Money" user, where Registered Mobile Money users could also be labelled as "Active Mobile Money" users according to the following criteria:

- **Voice Only Users:** If the user did not make any Mobile Money transactions during the 3-month evaluation period.

[3]In practice, the cell tower is accurate within several hundred meters in urban areas, and tens of kilometers in rural regions. We do not observe the contents of text messages. CDR are only generated when an individual initiates a transaction on the network, so we do not observe, for instance, the individual's location when she is not using her phone.

[4]Our intent was to exactly align the training and evaluation periods across countries, but implementation constraints made this impossible.

- **Registered Mobile Money User:** If the user made one or more Mobile Money transactions during the evaluation period.
- **Active Mobile Money User:** If the user made at least one Mobile Money transaction in each month of the evaluation period. Note that all Active Mobile Money users are also Registered Mobile Money Users.

# 3. FEATURE ENGINEERING WITH DETERMINISTIC FINITE AUTOMATA

## 3.1 Approach

As highlighted in the introduction, the CDR contain a wealth of latent information about how people communicate, with whom they interact, the locations they visit, and many other social and behavioral characteristics. Our eventual goal is to leverage this information to better understand why people use Mobile Money, and to develop a predictive model of Mobile Money adoption. However, the raw CDR are not natural inputs to most machine learning algorithms, and interpretable metrics must first be derived from the CDR before inferences can be made.

In the prior literature, the vast majority of studies take a rather ad hoc approach to constructing interpretable metrics ("features") from the phone data. The most common approach is to hand-craft a small number of features that correspond to some intuition of the researcher. For instance, [10] focus on 5 topological properties of the static social network; [17] use two metrics that quantify airtime purchases; and [14] construct 6 measures of physical mobility. Even the more ambitious approaches, such as [28] and [5], which respectively use 350 and 180 CDR-based metrics, employ a large number of idiosyncratic rules to determine which features should be considered by the learning algorithm. These approaches have the advantage of producing metrics that are convenient to interpret, but they may systematically overlook non-intuitive features, mis-attribute relationships (if, for instance, feature A is weakly correlated with the target variable only because an omitted feature B is strongly correlated with both A and the target variable), or fail to maximize the predictive power of a classifier that would perform better with a more comprehensive set of features.

Our approach is different. We develop a method for feature engineering from transactional data that is designed to construct a large and comprehensive set of features from a small number of recursive operations. While the application is to CDR, we believe this method could be used to engineer features from a more diverse class of data including IP logs, social media data, and financial transaction records.

## 3.2 Deterministic Finite Automaton

We employ a deterministic finite automaton (DFA), a model of computation from automata theory also referred to as a deterministic finite state machine, to formalize the feature generation process [22]. DFA's are typically used in more formal settings to determine whether an expression can be computed, to design circuits, or to operate simple devices. In the abstract, however, DFA's simply define a sequence of legal operations. We appropriate this concept to specify a set of legal operations that can be recursively applied to raw transactional data in order to produce valid features.

### *Example*

As an example, say we are interested in constructing a feature for each individual $i$ that corresponds to, "the variance in the average duration of outgoing calls made by $i$ on different days of the week." We allow for the construction of this feature through the following set of recursive rules:

1. filter *outgoing calls*
2. filter transactions *initiated by i*
3. group by *day of week*
4. focus on *call duration*
5. aggregate by *average* (duration per day of week)
6. aggregate using *variance* (over average daily durations)

By using different filter criteria (or difference group-by and aggregation operations), by adding and removing rules, or by applying the rules in a different order, we produce different features. It is important to note, however, that not all combinations of operations are valid. For instance, it does not make sense to take the variance of a categorical variable (such as `recipientID`), nor does it make sense to group by day of week if a day of week filter has already been applied. The power of the DFA is that it allows us to formalize the set of valid features using a relatively parsimonious specification.

### *Formalization*

The DFA we use to generate features from CDR is shown in Figure 4. In the figure, each circle represents a state, and for convenience we note the data structure expected for each state inside the circle. Valid features are constructed through traversals of the state machine, which start at the start state (`q0`) and end at the end state (`q3`) and follow only legal transitions between states (denoted by arrows). For instance, the feature described above would begin with the full CDR in `q0`, filter outgoing calls and return to `q0`, map (group by) "ego" $i$ and proceed to `q1`, map by day of week to `q4`, select duration and proceed to `q5`, reduce (aggregate) by average - this produces average call duration for each day of week for each $i$ - and proceed to `q2`, aggregate by variance and exit at `q3`.

Formally, the DFA is specified by:

- **Legal states:** `Q = {q0,q1,q2,q3,q4,q5,q6,q7}`
- **Start State:** `q0` $\in$ `Q`
- **End State:** `q3` $\in$ `Q`
- **Alphabet:** $\Sigma =$ {`CDRs, Ego-CDRs, Ego-f-CDRs, Ego-Values, Ego-value`}
- **Transition Functions** ($\delta : Q \times \Sigma \rightarrow Q$)**:** `filter, map, select, reduce`. The set of legal transitions is described in greater detail in Appendix A

In total, there are several thousand valid traversals of the DFA, each of which produces a different feature. Together, the resulting set of features covers almost all of the hand-crafted metrics used in prior work, as well as many, many more.

An additional advantage of the DFA is that it can be efficiently and elegantly implemented. Pseudo-code is provided in Algorithm 1, and the implementation in Spark Python is available from the author's website.[5]

[5]http://www.jblumenstock.com

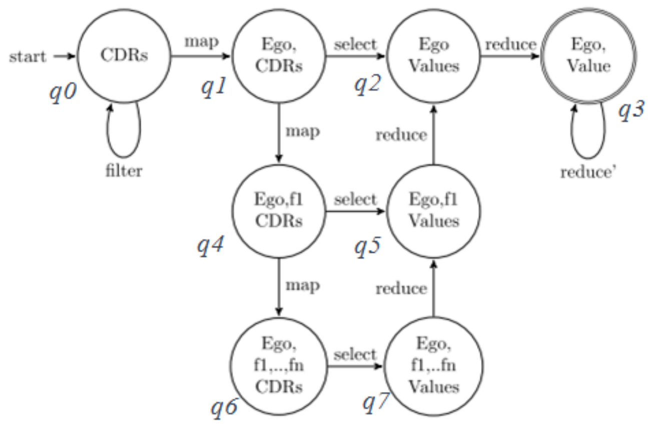


**Figure 4: Deterministic Finite Automaton**

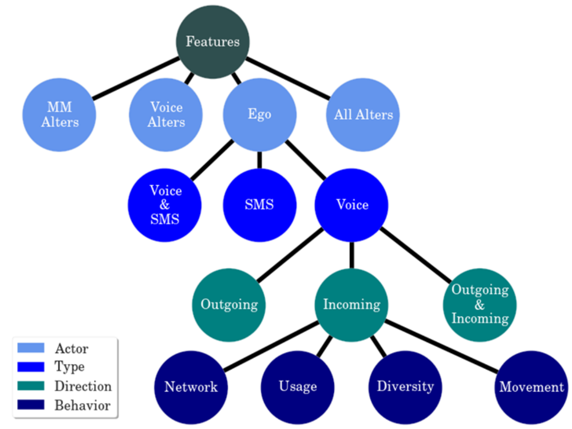


**Figure 5: Tree-based feature classification**

## 3.3 Feature classification and tree structure

The DFA is a convenient abstraction for generating a very large number of features from a small set of rules. To interpret the set of features produced by the DFA, we label each generated feature with interpretable tags. These tags are determined by the path taken through the automaton, and indicate whether each feature captures information on, for example, incoming vs. outgoing communications, calls vs. text messages, variance vs. volume, and so on. Specifically, we map each feature onto the tree structure shown in Figure 5, which is designed to encapsulate the substantive behavior captured by each feature. Each level of the tree corresponds to a different partition of the feature space:

1. **Actor:** Whether the feature relates to activity of the individual $i$ (the "ego"), or the activity of $i$'s first degree network of connections (the "alters"). We separately look at $i$'s full set of alters, the set who have previously used Mobile Money, and the set that have never used Mobile Money. A simple aggregation operation (such as `mean` or `SD`) is applied to the alter network to produce a feature for $i$.
2. **Type:** Whether the feature relates to phone calls or text messages (SMS).
3. **Direction:** Whether the feature relates to incoming (e.g., call received) or outgoing (e.g., call placed) activity.
4. **Behavior:** Whether the feature relates to movement (e.g., number of unique cell towers used), network structure (e.g., number of unique contacts in network), phone usage (e.g., number of calls made), or Diversity (e.g., geographic spread of social network). This tag is determined by the data type of the field over which aggregated is performed (e.g., continuous vs. discrete data) and the actual statistical function used in aggregation (`count, unique, min, max, mean, median, SD, variance, radius of gyration`)

Figure 5 is simplified to show only a single expansion along the ego-voice-incoming path. In practice, all nodes on a given level can be expanded analagously to the path shown in Figure 5. For instance, the example feature described in Section 3.2 would be a leaf on the branch of ego-voice-outgoing-usage.

**Algorithm 1:** Feature Generation Algorithm

**Data:** cdr,Call Detail Records of all users
**Data:** opmap,Dictionary of possible operations
*cdrtypes* ← [Voice, SMS, Voice and SMS]
*direction* ← [In, Out , In plus Out]
*featuresArray* ← []
**Result:** Features

Step 1: Perform reduce by grouping only on ego
**foreach** *type, dir1 in cdrtypes,direction* **do**
  *filteredCDR* ← *cdr*.filter(*type*,*dir*1)
  **foreach** *field in cdr* **do**
    *groupeddata* ←
     *filteredCDR*.map([*ego*]+[combinations(*field*)])
    **foreach** *op in opmap[field]* **do**
      *reduceddata* ← *groupeddata*.reduce(*op*)
       insert *reduceddata* in *featuresArray*
      *reducedata2* ←
       *reduceddata*.map(*ego*,*alters*).reduce(*op*)
       insert *reduceddata2* in *featuresArray*
    **end**
  **end**
**end**

# 4. MODELS AND METHODS

The DFA-based process of feature engineering described above generates thousands of features that quantify patterns of mobile phone use. Armed with these features, our goals are to (a) use these features to understand the determinants of Mobile Money use, (b) build a predictive model that can be used to identify likely adopters, and (c) determine the extent to which models and features from one context can generalize to another.

## 4.1 Experimental Design

To facilitate our supervised learning experiments, we drew a stratified random sample of 10,000 subscribers from each of

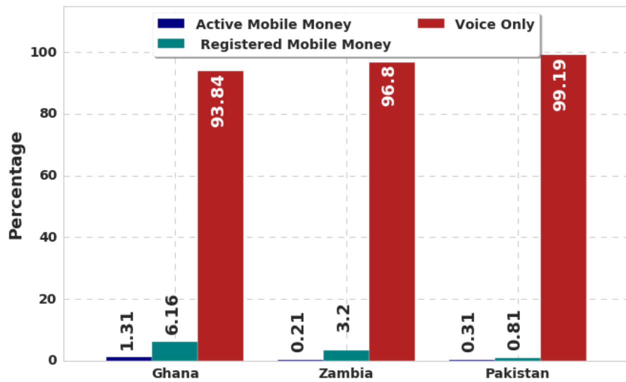


**Figure 6: Distribution of user types by country**

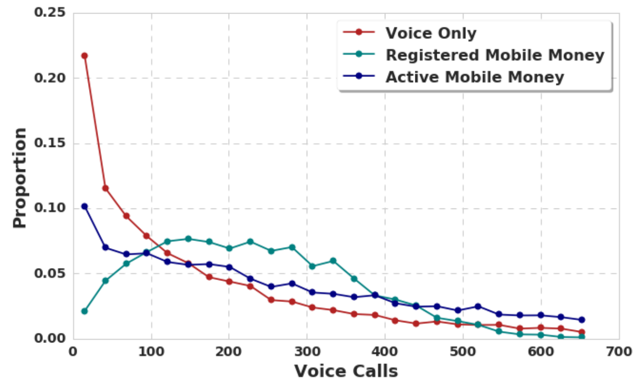


**Figure 7: Distribution of calls per subscriber, Ghana**

the three categories (Voice Only, Registered Mobile Money, Active Mobile Money) from each country. We elected to draw a balanced sample since, as can be seen in Figure 6, the vast majority of subscribers in each country fall into the "Voice Only" category.[6]

## 4.2 Classification and Model Selection

We then use a variety of supervised learning algorithms to tackle two classification tasks. First, we seek to differentiate between Voice Only subscribers and Registered Mobile Money Users (one or more Mobile Money transactions); second, we attempt to differentiate between Voice Only and Active Mobile Money users (at least one transaction per month). In all cases, we report the average accuracy across testing sets from 10-fold cross validation.

Since our data has a large number of features relative to observations, we focus on learners that are robust to overfitting, such as regularized and elastic net logistic regression [30], gradient boosting [15], and Extremely Randomized Trees [16]. Performance was comparable across these classifiers, although as expected these methods generally performed better than unregularized alternatives. To streamline the analysis that follows, we report only the results from gradient boosting, which outperformed the other classifiers by a small margin.

## 4.3 Feature selection and importance

To understand which CDR-based features are related to Mobile Money use, we calculate two metrics:

1. **(Unconditional) AUC**: We run a (cross-validated) bivariate logistic regression of the response variable (one of the above definitions of Mobile Money use) on each feature separately. The provides an indication of the unconditional correlation between each feature and the response variable.
2. **(Conditional) Normalized feature importance**: We calculate the importance of each feature to the final gradient boosting classifier. As we are primarily interested in the *relative* importance, the set of feature importances is standardized to be comparable across countries and classification tasks. Following [15], we denote the relative influence of feature $x_j$ in tree $T$ as

$$\hat{I}_j^2(T) = \sum_{t=1}^{J-1} \hat{i}_t^2 1(v_t = j)$$

where $\hat{i}^2$ is the improvement in (squared) error achieved by splitting feature $v_t$ at node $t$, summed over all non-terminal nodes $J \in T$. In a collection of gradient-boosted trees, the average feature importance $\bar{I}_j$ is the arithmetic mean of $\hat{I}_j^2(T)$ across all trees, and the *normalized feature importance* is the z-score obtained by subtracting the mean (of all $\bar{I}_j$) and dividing by the standard deviation (of all $\bar{I}_j$) for each $\bar{I}_j$.

[6]To protect the commercial interests of the operators, we show only the fraction of users of each type, rather than the raw numbers, which are in the millions.

# 5. RESULTS

## 5.1 Determinants of Mobile Money Adoption

The DFA described in Section 3 produces roughly 3,000 unique features. As one example, Figure 7 shows the distribution of total calls made per subscriber in Ghana, for each of the three subscriber types. There are clear differences between the three user types in this distribution, with Voice Only users making the fewest calls, Registered Mobile Money users concentrated in the range from 100-300 calls (in the 10-day training period), and Active Mobile Money users more evenly distributed across the full range from 100-700 calls.

The distribution of unconditional AUC values for each of the 3,000 features is shown in Figure 8 (left panel), using Ghana as a test case. To construct this figure, we use the feature classification schema from Figure 5 to label each feature with four tags corresponding to the Actor, Type, Direction, and Behavior of the feature. Each violin plot then shows the distribution of AUC values for all features of a given type - such as all "ego" features, or all "movement" features. The left (blue) half of each violin plot indicates the distribution of AUC values for features when discriminating between Voice Only and Registered Mobile Money; the right (red) half shows the distribution when discriminating between Voice Only and Active Mobile Money.

While a large number of features have AUC values near 0.5, indicating they contain little information about the distinction between Voice Only and Mobile Money users, several noteworthy patterns emerge. First, when feature types are defined by the coarse classification tree in Figure 5, no

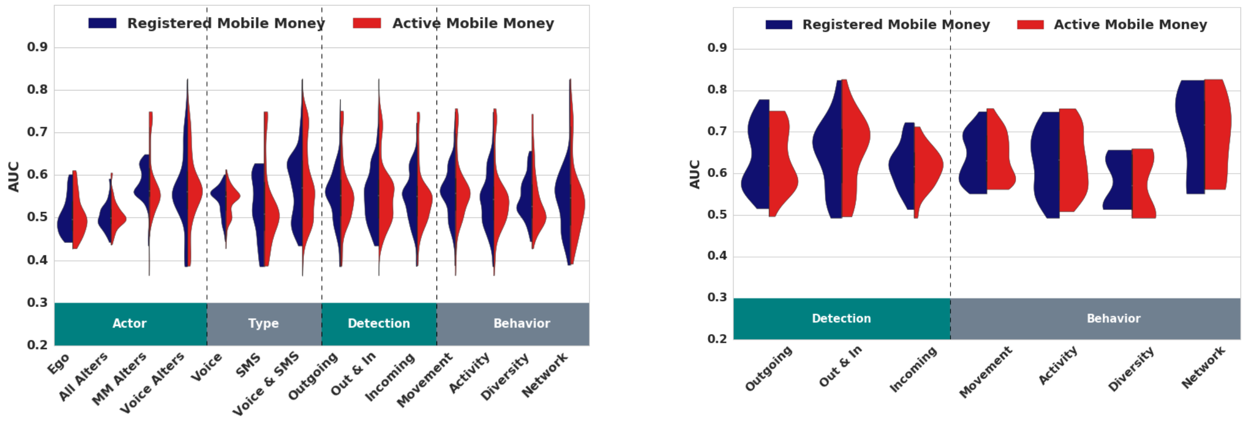


**Figure 8: Distribution of AUC values for each feature category. Left figure shows all features in Ghana; Right figure shows the subset of features in Ghana where Actor='Voice Alters' and Type='All'**

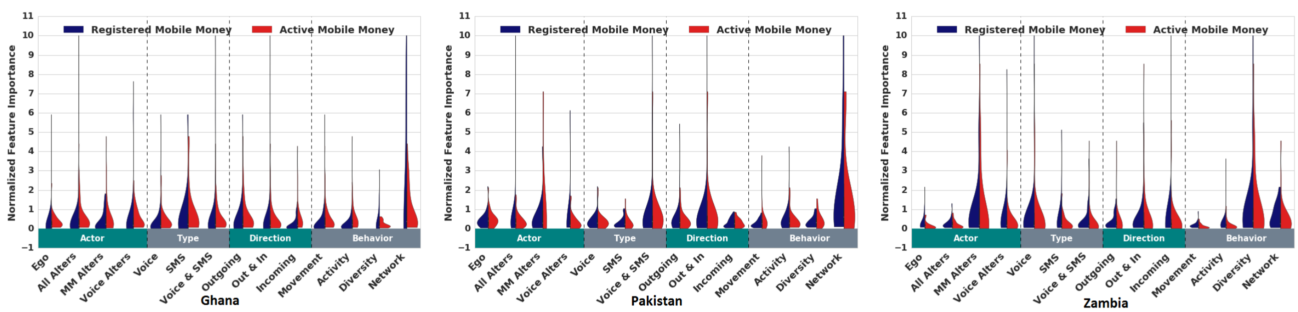


**Figure 9: Normalized feature importance**

single type dominates; rather, most types of features have a large number of uninformative features and a small number of highly predictive features with AUC≥0.75. At the same time, feature classes do matter. The right panel of Figure 5 shows the distribution of AUC values for the subset of features where Actor="Voice Alters" and Type="All", a subset that are generally more predictive of Mobile Money use in Ghana. Here, the range of AUC values is significantly higher than in the full set of features, and some sub-classes such as "Network" have uniformly high predictive power.[7] Finally, the usefulness of each class of features depends on whether the goal is to identify Registered or Active Mobile Money users. For instance, the "MM Alter" features, which capture information about the characteristics of $i$'s network who have previously adopted Mobile Money, are bimodally distributed and on average more useful in predicting Registered users than Active users. However, that same class contains a small number of features that are extremely good predictors of Active Mobile Money use.

A similar approach is taken to construct Figure 9, except here we show the distribution of normalized feature importance values obtained through gradient boosting. The difference between the values in Figure 9 and Figure 8 is that the former are conditional on all features present in the final classification model, which includes several hundred features, whereas the latter are unconditional, i.e., they indicate performance in a univariate model with no other features. As in the unconditional ranking, each class of features in the conditional ranking contains a mass of features with low predictive power, but closer inspection reveals interpretable patterns.

[7]This particular class, where Actor="Voice Alters," Type="All," and Behavior="Network", corresponds to information about the network structure of $i$'s network; in other words, 2nd degree properties of $i$'s network.

Perhaps most striking in Figure 9 are the differences between countries in the relative importance of each class of features. For instance, we see that in Ghana and Pakistan the "Network" features are in general more important to the classification model than the other types of Behavior, whereas in Zambia "Diversity" is most important. Zambia is also unique in the higher importance placed on voice calls relative to SMS activity, and in the fact that more signal exists in incoming calls than in outgoing calls. As we discuss below, these cross-country differences imply that models trained in one context may not generalize well to others.

## 5.2 Predicting Mobile Money Use

As discussed in Section 4.2, we test the ability of several supervised classification models to discriminate between Voice Only and Mobile Money users, using the CDR-based features constructed from the DFA. Cross-validated results from gradient boosting are reported separately for each country in Figure 10. The best-performing models included several hundred features, but in practice there was little difference in performance between models in the range of 50-1000 features. We also include results from a baseline classifier, which uses the same model trained on a single "intuitive"

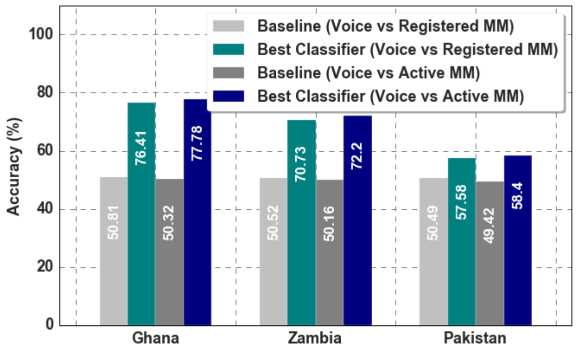


**Figure 10: Accuracy in identifying Mobile Money users within each country, using gradient boosting**

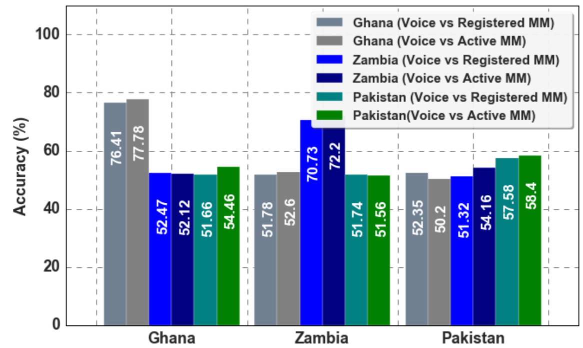


**Figure 11: Accuracy when model is trained in one country and evaluated in another**

feature – the total number of outgoing calls made by the subscriber, which is the feature shown in Figure 7.

In each country, the DFA-based classifier significantly outperforms the naive baseline, and in all countries, we achieve marginally better success in identifying Active Mobile Money users than Registered Mobile Money users. Across countries, however, there is a great degree of variability in classifier performance, with classification accuracy between 71% and 78% in Ghana and Zambia, but only 58%-59% in Pakistan. We discuss several possible explanations for these results in Section 6.

## 5.3 "Transfer Learning"

In the proceeding analysis, we have been careful to standardize the methods and analysis performed across all three countries. In each instance, we use the exact same source data, DFA specification, classification algorithm, experimental sample size, and so forth. In some cases, this meant that we knowingly discarded data that might have improved the performance of the classifier in a single country. For example, in some countries we had several months of CDR that could be used for training, additional fields in the CDR metadata, or much larger samples of Mobile Money users available for training and cross-validation. However, our approach reduced everything to the lowest common denominator in order to maintain comparability across contexts.

A key advantage of this approach is that it makes it possible to answer a question that has been elusive in prior studies of the adoption of new technologies in developing countries: *Do the behavioral determinants of adoption identified in one context generalize to another?* Based on the analysis we have performed, our short answer to this question appears to be, "No."

Figure 11 shows the performance of a classifier trained in one country and evaluated in another. Thus, the first set of six bars shows that the classifiers trained in Ghana perform well in Ghana (the first two grey bars), essentially replicating the results in Figure 10. However, that same Ghana model does quite poorly when it is evaluated in Zambia (the next two blue bars) and Pakistan (the final two green bars). While it is almost certain that a more sophisticated approach to transductive transfer learning would perform better [25], the naive application of a model out of context is quite ineffective. We return to these ideas in the discussion that follows.

# 6. INTERPRETATION AND DISCUSSION

Taken in the broader context of research into the determinants of Mobile Money adoption and use, the preceding results uncover several unexpected patterns. Superficially, it is not surprising that CDR-based metrics can be used to construct classifiers that predict Mobile Money use, though to our knowledge this is the first study to publish performance metrics that can serve as benchmarks in future work in this area. However, in our analysis we were surprised to find that in a given country, the supervised model was only marginally better able to identify Active Mobile Money users (who make at least one transaction per month) than Registered Mobile Money users (who make at least one transaction ever). By contrast, our expectation was that active users, who are quite rare in all three countries, would have distinct patterns of phone use that would make them easier to detect. Since most policymakers agree that true financial inclusion requires active use, this remains an open topic for future work.

Also interesting are the differences in performance of the same modeling approach applied in different contexts (Figure 10). Most striking here is the relatively poor performance in Pakistan, where the 18% improvement over the baseline is dwarfed by the 55% improvement over the baseline achieved in Ghana.[8] At face value, this finding implies that Mobile Money users in Pakistan are very similar to non-Mobile Money users, or at least that the two groups have similar patterns of mobile phone user. However, looking more carefully at the data, we believe this may also in part be an artifact of the "one size fits all" approach we have taken to standardizing definitions and methods across countries. In particular, there is one type of Mobile Money transaction that is extremely common in Pakistan, which allows a subscriber to add prepaid phone credit to her phone account

[8] This result is also unexpected, given the evidence in Figure 9, which indicates that Ghana and Pakistan have very similar profiles in terms of relative feature importance. If anything, this figure might lead one to suspect that Zambia would be an outlier in the analysis, since Zambia's feature profile is distinct from the other two countries.

using Mobile Money. Anecdotally, it is common practice in Pakistan for the retailers of phone credit to perform this Mobile Money transaction on behalf of the subscriber. Thus, a subscriber might appear to be using Mobile Money, though in practice she was not responsible for the transaction. This potential source of bias highlights the brittle nature of the cross-country analysis, which in its current form does not allow for country-specific adaptation.

Perhaps most importantly, our results suggest that across different countries and cultures in the developing world, no single set of behavioral features is likely to consistently predict Mobile Money adoption and use. This is most clearly evident in Figure 11, which shows that a classifier trained in one country performs very poorly when tested in another country. But the same conclusion may also be drawn from Figure 9, where we see that the same model, when trained in different countries, selects different features and attaches different weights to to those selected features. In results not shown, we further inspect the list of top-ranked features for each country, using both (unconditional) AUC and (conditional) normalized feature importance, and note very few features that appear consistently across countries. However, even though there may not be a "golden" list of features that always predict Mobile Money use, we are optimistic that more generalized insights can be extracted from one context and applied in another. In ongoing work, we are exploring methods for transfer learning that may strike this balance.

More concretely, over the past several months our partner in Ghana has been using the methods we describe to generate "Adoption Scores" that indicate the likelihood that any given mobile subscriber will adopt and use Mobile Money. They recently reported that when using these scores to target promotions, response rates were roughly 30% higher than promotions targeted with traditional methods. Such estimates are notoriously unreliable and subject to many possible sources of bias, but their optimism provides an indication of the potential for this line of research. At the same time, it should be noted that if the end goal is to increase financial inclusion of the poor, further methodological innovation is needed beyond what identifies the "low hanging fruit" subscribers whose behavior indicates that they are likely to adopt of their own volition.

# 7. CONCLUSION

In this paper, we present a new approach to feature engineering that uses deterministic finite automata to construct a very large number of features from a concise set of rules. In applying this technique to mobile phone data from Ghana, Pakistan, and Zambia, we show that the resultant metrics correlate with, and can be used to predict, both active and passive Mobile Money use in three very different contexts. In so doing, we discover several previously undocumented patterns related to the adoption and use of Mobile Money. Superficially, the analysis makes it possible to highlight specific correlates of Mobile Money use, such as the relative importance of network structure in Ghana and Pakistan, and the relative importance of geographic diversity in Zambia. More fundamentally, the results provide insight into the extent to which standard predictive models can generalize across contexts. Here, it is clear that each population has a unique signature in terms of what metrics are good predictors of adoption, and as a result, models trained in one location do not perform well in another. Retraining the model helps, but does not solve, the underlying issue. Despite the fact that the data structures, experimental design, and Mobile Money products are nearly identical in the three countries, the performance of each country-specific model varies greatly.

# 8. ACKNOWLEDGMENTS

We are grateful to Sven Harten of the International Finance Corporation for supporting this work and providing thoughtful feedback throughout the research process. Gabriel Cadamuro and Robert On were instrumental in developing an earlier version of the DFA-based approach to feature engineering.

# 9. REFERENCES

[1] J. Aker, R. Boumnijel, A. McClelland, and N. Tierney. How do electronic transfers compare? Evidence from a mobile money cash transfer experiment in Niger. *CGD Working Paper 268*, Washington, D.C., 2014.

[2] J. Blumenstock, G. Cadamuro, and R. On. Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076, 2015.

[3] J. E. Blumenstock, M. Callen, T. Ghani, and L. Koepke. Promises and pitfalls of mobile money in afghanistan: evidence from a randomized control trial. In *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development*, page 15. ACM, 2015.

[4] J. E. Blumenstock, N. Eagle, and M. Fafchamps. Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters. *Journal of Development Economics (forthcoming)*, 2016.

[5] CGAP. The Power of Social Networks to Drive Mobile Money Adoption. Technical report, 2013.

[6] G. Chen and S. Rasmussen. bKash Bangladesh: A Fast Start for Mobile Financial Services. Technical report, 2014.

[7] A. Demirguc-Kunt, L. F. Klapper, D. Singer, and P. Van Oudheusden. The global findex database 2014: measuring financial inclusion around the world. *World Bank Policy Research Working Paper*, (7255), 2015.

[8] A. Dermish, C. Kneiding, P. Leishman, and I. Mas. Branchless and mobile banking solutions for the poor: A survey of the literature. *Innovations: Technology, Governance, Globalization*, 6(4):81–98, Oct. 2011.

[9] S. Di Castri and L. Gidvani. Enabling Mobile Money Policies in Tanzania: A "test and learn" approach to enabling market-led digital financial services. Technical report, GSMA, Feb. 2014.

[10] Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla. Inferring user demographics and social strategies in mobile social networks. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 15–24. ACM, 2014.

[11] K. Donovan. Mobile money for financial inclusion. Information and Communications for Development, pages 61–73. The World Bank, 2012.

[12] A. S. Etim. Mobile banking and mobile money adoption for financial inclusion. *Research in Business and Economics Journal*, 9:1, 2014.

[13] V. Frias-Martinez, E. Frias-Martinez, and N. Oliver. A gender-centric analysis of calling behavior in a developing economy using call detail records. 2010.

[14] V. Frias-Martinez, J. Virseda, and E. Frias-Martinez. Socio-Economic Levels and Human Mobility. *Jourhnal of Information Technology for Development*, pages 1–16, Feb. 2012.

[15] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
[16] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Mach. Learn.*, 63(1):3–42, Apr. 2006.
[17] T. Gutierrez, G. Krings, and V. D. Blondel. Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets. arXiv e-print 1309.4496, Sept. 2013.
[18] W. Jack and T. Suri. Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution. *American Economic Review*, 104(1):183–223, 2014.
[19] M. R. Khan, J. Manoj, A. Singh, and J. Blumenstock. Behavioral modeling for churn prediction: Early indicators and accurate predictors of custom defection and loyalty. In *2015 IEEE International Congress on BigData*, pages 677–680. IEEE, 2015.
[20] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5, 2007.
[21] I. Mas and D. Radcliffe. Scaling mobile money. *Journal of Payments Strategy & Systems*, 5(3):298–315, 2011.
[22] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, Dec. 1943.
[23] I. Medhi, A. Ratan, and K. Toyama. Mobile-banking adoption and usage by low-literate, low-income users in the developing world. In *Internationalization, Design and Global Development*, pages 485–494. Springer, 2009.
[24] O. Morawczynski. Exploring the usage and impact of "transformational" mobile financial services: the case of m-PESA in kenya. *Journal of Eastern African Studies*, 3(3):509–525, Oct. 2009.
[25] S. J. Pan, I. Tsang, J. Kwok, and Q. Yang. Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, Feb. 2011.
[26] Safaricom. FY14 Presentation, 2014.
[27] C. Scharwatt, A. Katakam, J. Frydrych, A. Murphy, and N. Naghavi. State of the Industry: Mobile Financial Services for the Unbanked. Technical report, GSMA, 2014.
[28] P. Sundsøy, J. Bjelland, A. M. Iqbal, Y.-A. de Montjoye, et al. Big data-driven marketing: How machine learning outperforms marketers' gut-feeling. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 367–374. Springer, 2014.
[29] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16):5962–5966, 2012.
[30] H. Zou and T. Hastie. Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.

# APPENDIX

## A. DFA TRANSITION DETAILS

The Deterministic Finite Automaton described in Section A is defined in part by a set of transition functions $\delta : Q \times \Sigma \to Q$ that define the legal transitions between states $Q$:

- *Filter*: $Q \times CDR \to CDR$
  Operations filter the data on the basis of the type of communication (voice, sms) or the direction the activity (incoming, outgoing).
- *Map*: $Q \times CDR \to (field, CDR)$
  Operations are similar to 'group by' operations, which group the input CDR according to the value of a field in the CDR. All data are initially mapped by ego $i$, creating $N$ subsets of CDR, one corresponding to each individual. Other common *map* criteria include alter, location, and day-of-week.
- *Select*: $Q \times (field^*, CDR) \to (field^*, fields)$
  Operations extract a field (or column) from the CDR. Common select criteria are duration, location, and a constant 1 used to count transactions.
- *Reduce*: $Q \times (field^*, fields) \to (field^*, value)$
  Operations 'aggregate' a set of values into a single value. Valid operations include unique, count, average, standard deviation, radius of gyration, and entropy. Whether a *reduce* operation is legal depends on the data type of the field to which it is being applied. *Reduce'* is a special operation that aggregates values across all of an ego's alters, creating a feature that represents the {average, standard deviation} of the individual's first-degree network.