



GENDER Impact
Platform

CGIAR GENDER Impact Platform · Working Paper #023

DECEMBER 2024

Can machine-learning models predict gendered labor statistics using mobile phone and geospatial data?

By Greg Seymour, Lendie Follett, Heath Henderson,
Nathaniel Ferguson



This publication is copyrighted by the International Livestock Research Institute (ILRI). It is licensed for use under the Creative Commons Attribution 4.0 International License. To view this license, visit <https://creativecommons.org/licenses/by/4.0>.



Unless otherwise noted, you are free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform and build upon the material) for any purpose, even commercially, under the following conditions:

④ **ATTRIBUTION.** The work must be attributed when used in other works, but not in any way that suggests endorsement of those works by ILRI or the author(s).

ISBN: 92-9146-838-2

Citation: Seymour, Greg, Follett, Lendie, Henderson, Heath, and Nathaniel Ferguson. 2024. *Can machine-learning models predict gendered labor statistics using mobile phone and geospatial data?* CGIAR GENDER Impact Platform Working Paper #023. Nairobi, Kenya: CGIAR GENDER Impact Platform.

ACKNOWLEDGMENTS

This work was supported by the CGIAR GENDER Impact Platform and Digital Innovation Initiative. This work would not be possible without the commitment and collaboration of all partners. We thank Frederic Pivetta and Brian King for their input in the development of the project proposal. We thank Omar Seidu, David Bessah, Gloria Akoto-Bamfo, Anthony Oduro-Denkyirah, Victor Owusu, Rachel Bowers, Laurent Smeets and Abena Oduro for their input in the development of the phone survey. We thank the Ghana Statistical Service enumeration team for their hard work and dedication during the implementation of the phone survey. We thank Nancy Lubalo and Denys Sementsov for preparing and processing the mobile phone data and implementing the machine-learning model. We thank Astrid van Lierde for her indispensable support in managing the project throughout all phases. We thank two anonymous reviewers for their feedback that helped to strengthen the paper.

We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund: <https://www.cgiar.org/funders/>.

COVER PHOTO: Mapping workshop in Nyangania, Kassena Nankana District, Ghana. Credit: Axel Fassio/CIFOR.

ABOUT CGIAR GENDER IMPACT PLATFORM

Generating Evidence and New Directions for Equitable Results (GENDER) is CGIAR's impact platform designed to put equality and inclusion at the forefront of global agricultural research for development. The Platform is transforming the way gender research is done, both within and beyond CGIAR, to kick-start a process of genuine change toward greater gender equality and better lives for smallholder farmers everywhere. gender.cgiar.org

DISCLAIMER

This working paper has been internally peer-reviewed and the opinions expressed herein reflect those of the authors, not necessarily those of the CGIAR GENDER Impact Platform.

CONTACT

Greg Seymour
Survey Statistician, US Census Bureau
Email: gregory.t.seymour@census.gov

CGIAR GENDER Impact Platform Working Paper #023

DECEMBER 2024

Can machine-learning models predict gendered labor statistics using mobile phone and geospatial data?

Greg Seymour

US Census Bureau

gregory.t.seymour@census.gov

Lendie Follett

Drake University

lendie.follett@drake.edu

Heath Henderson

Drake University

heath.henderson@drake.edu

Nathaniel Ferguson

International Food Policy Research Institute

n.ferguson@cgiar.org

TABLE OF CONTENTS

Abstract	1
1. Introduction	2
2. Related work	4
3. Data	5
<i>3.1 Phone survey</i>	6
<i>3.2 Digital trace data</i>	8
<i>3.3 Geospatial data</i>	9
4. Methodology	11
<i>4.1 Prediction models</i>	11
<i>4.2 Estimation of machine-learning models</i>	13
<i>4.3 Cross-validation</i>	14
5. Results	15
<i>5.1 Cross-validation of predictions</i>	15
<i>5.2 Prediction mapping</i>	17
<i>5.3 External validation</i>	21
6. Discussion and conclusions	22
References	25
Appendix 1. Phone survey questionnaire	28
Appendix 2. Mobile phone usage features used in the analysis	31
Appendix 3. Summary information for remote sensing–GIS datasets used for covariate selection	33

Abstract

High-quality data on rural women's and men's labor is imperative for tracking progress on gender equality and women's empowerment, and for evaluating development interventions aimed at these outcomes. Yet, there remains a general lack of sex-disaggregated data on unpaid care and domestic work, earnings, employment and entrepreneurship. Researchers are increasingly looking to digital technologies, such as mobile phones, as an emerging data source with significant potential for closing gender data gaps. In this paper, we attempt to use mobile phone data and machine-learning models to predict gendered labor-market indicators for a large sample of mobile phone users in Ghana. Although our models predict mobile phone subscribers' sex with reasonable accuracy, they predict women's and men's labor-market outcomes only slightly better than random guessing. The models' mixed results may be partly attributed to noisiness in the data due to disruptions in mobile phone- and employment-related behaviors caused by COVID-19. Our results also point to potential methodological limitations in using machine-learning methods and mobile phone data to estimate gendered labor-market indicators, and more generally suggest that we should proceed cautiously when thinking about leveraging digital technologies and machine learning to close data gaps. We conclude the paper with several recommendations for how the methodology might be refined in future work.

Keywords: *gender, labor, mobile phones, machine learning, gender statistics, big data*

1. Introduction

Accurate, sex-disaggregated data on women's and men's livelihoods are critical for research and policy. Such data provides the foundation for global efforts to track progress on gender equality and inform the design of policies to overcome gender-based constraints and transform restrictive gender norms. Although the quantity and quality of data available in low- and middle-income countries (LMICs) have improved in recent years, data on women and girls are still lacking in many LMICs, arising both from a failure to prioritize gender equality in data collection and from a lack of resources to collect and analyze data (Buvinić, Furst-Nichols and Koolwal 2014; Lopes and Bailur 2018).

Some of the most pressing gender-data gaps relate to unpaid care and domestic work, earnings, informal employment, and entrepreneurship (Data2x and Grantham 2020). Labor-force and household surveys, which are the primary sources of individual-level information for these data, can be slow to fill in data gaps, especially in LMICs where they are infrequently implemented (Desiere and Costa 2019). Moreover, nationally representative labor-force and household surveys often lack spatial granularity, impeding governments' ability to monitor fluctuations in labor-market vulnerabilities at subnational levels. There is thus a critical need to identify new methodologies to address existing data gaps.

Big data based on digital technologies, such as mobile phones, are an emerging source with significant potential (Lopes and Bailur 2018). Digital trace data, including call-detail records (CDRs), are produced every time a mobile phone is used, including when people make calls, send text messages, browse the internet, or pay for goods and services using mobile money.¹ Linked with passively generated metadata such as timestamps and cell tower location, digital trace data can be used to generate variables (or features) related to a variety of individual behaviors, including calling patterns, physical mobility, and social networks.² Such data typically cover substantially wider populations, often more cost-effectively and at greater spatiotemporal resolution than traditional labor-force or household surveys.

A growing body of literature has proposed using modern machine-learning methods coupled with digital trace data to address data gaps. This approach typically consists of three steps: (1) fit a variety of machine-learning models to predict an outcome of interest using digital trace data linked with existing survey data for a sample of mobile phone users, (2) select the best performing model, and (3) use the estimated model to predict the outcome for the entire population of mobile phone users. Blumenstock et al. (2015), for example, merged a survey-based wealth index with CDRs from Rwanda's largest mobile phone operator, generated several thousand features from the CDRs, used a machine-learning algorithm to eliminate irrelevant features and select a parsimonious model, and then applied their model to predict the wealth of all subscribers to the mobile phone network. Similar methods have been used to predict social networks, movement patterns, shock coping, and poverty status.³

In this paper, we examine whether machine learning combined with rich digital trace data can be used to address gender data gaps associated with key labor-market indicators—specifically employment, unemployment and underemployment. Notably, we do not predict unpaid work and domestic activities. Though the machine-learning approach has

1 Mobile money services, the use of which has spread quickly throughout lower-income countries in recent years, allow customers to perform financial transactions outside of formal banks (Lavelle-Hill et al. 2022).

2 This process is generally referred to as *feature extraction* or *feature engineering*.

3 Leading examples include Eagle, Macy and Claxton (2010) or Gutierrez, Krings and Blondel (2013) on social networks; Blumenstock (2012) or Lavelle-Hill et al. (2022) on movement patterns; Blumenstock, Eagle and Fafchamps (2016) or Lu et al. (2016) on shock coping; and Steele et al. (2017) or Aiken et al. (2022) on poverty status.

been shown to perform adequately in many nongendered applications, there are several reasons to expect limited performance in terms of addressing gender data gaps. First, model performance depends on the capacity of digital trace data to predict outcomes for both women and men, but such data may be less predictive of women's outcomes. Rural women's mobile phone usage, for example, may be relatively less predictive of employment-related activities because in LMICs, they are less likely to work outside the home, less likely to carry their phone outside the home, and less likely to use their phones for work (see [Section 3.1](#)).

Second, the performance of the machine-learning approach depends heavily on the representativeness of the digital trace data for both women and men. While digital trace data are generally unrepresentative of the underlying population, this issue may be especially pronounced for women because in low- and middle-income countries, they are less likely to have access to mobile phones than men (GSMA 2022).⁴ Additionally, some evidence suggests that women and men who own phones tend to be more similar than women and men in the general population—there may be systematic differences between phone and non-phone users that could lead prediction models to understate gender disparities. Blumenstock (2012) found limited differences in predicted migration and mobility patterns between women and men, which contrasts starkly with more direct evidence showing robust differences.

Finally, while the vast majority of existing machine-learning approaches focus on predicting a single outcome (e.g., wealth, income, or poverty), addressing gender data gaps using digital trace data requires predicting both the outcome of interest and gender. While data on gender is routinely collected during the purchase and registration of subscriber identity module (SIM) cards, this information is frequently unavailable to researchers due to data privacy restrictions, as is the case in our analysis. Even when available, the information is often inaccurate, as the actual user may differ from the registered user (Al-Zuabi, Jafar and Aljoumaa 2019). This is particularly true in low-income settings where men commonly register SIM cards for their wives and daughters, where agents often operate in very basic facilities and track registrations on paper, and where subscribers are unable to register in their own names due to lack of access to a legally valid ID (GSMA 2022). As a result, demographic information gathered at registration typically cannot be used reliably to distinguish male and female users. Consequently, estimating gendered economic indicators becomes a complex and challenging joint prediction problem. To determine whether the machine-learning approach is a viable route for addressing gender data gaps, we use a completely anonymized database of millions of mobile phone transactions from MTN Ghana (hereafter 'MTN'), the largest mobile phone provider in the country. We supplement this information with two additional data sources. First, we augment the mobile phone data with non-CDR, passively collected geospatial data (e.g., rainfall, temperature, travel times to cities, etc.) obtained from public databases and processed for all of Ghana. Second, we obtain our outcomes of interest from a phone survey conducted with a geographically stratified random sample of 8,066 MTN subscribers.⁵ From these survey data, we construct individual-level indicators of labor-market participation such as employment, unemployment, and underemployment—based as much as possible on International Labour Organization (ILO) (2013) standard definitions.

For each of the 8,066 survey respondents, we have information on labor-market participation, a complete record of their mobile phone activity over a six-month reference period, and geospatial data linked to these records based on the location of the cell tower that routed the majority of each subscriber's activity over those six months. We use this merged dataset to train various machine-learning algorithms to jointly predict the subscriber's sex and labor-market participation based on users' historical mobile phone usage patterns and the geospatial features associated with each subscriber's location. We further apply the model

4 While statistical reweighting techniques can be used to correct for selection biases, the success of these solutions varies depending on survey-specific factors (Pestre, Letouzé and Zagheni 2020; Brubaker, Kilic and Wollburg 2021; Gourlay et al. 2021; Ambel, McGee and Tsegay 2021).

5 Upon contacting subscribers, we obtained informed consent to merge their survey responses with their mobile phone transaction history. The phone survey solicited no personally identifying information; and included questions on mobile phone use, work, employment, and time use.

to the entire MTN subscriber base, using the predictions to develop spatially granular, gendered labor-market indicators for all of Ghana.

Our results show that addressing gender data gaps with machine learning is a formidable challenge. While our models can predict the sex of mobile phone subscribers with accuracy comparable to past studies (e.g., Slavchevska et al. 2021), their ability to predict labor-market outcomes is only slightly better than random guessing, and is worse than results reported in similar studies, most of which aim to predict poverty status. Our study, however, differs significantly from previous research. In part, this difference stems from the joint prediction problem we face in predicting both gender and labor outcomes. In addition, our data may be noisier: labor-market outcomes are inherently more prone to short-term fluctuations than poverty status. Furthermore, the reference period for both the labor and digital trace data used in our study coincides with a particularly tumultuous phase of the COVID-19 pandemic in Ghana, likely reflecting behavioral disruptions caused by changing COVID-19 restrictions during this time. Nevertheless, the mixed results obtained in our study point to potential limitations of using machine-learning methods and mobile phone data to estimate gendered labor-market indicators and similar outcomes.

2. Related work

While the raw data included in CDRs is limited to the occurrence, duration, and location of calls, texts and internet use, a wide variety of indicators can be obtained from this data through "feature extraction." Jahani et al. (2017) extracted over 1,447 features from CDRs, including temporal features such as the average duration of calls and text response rates, and spatial features such as radius of gyration—which measures phone users' movements relative to a central point. Sundsøy et al. (2017) extracted three categories of features: financial features such as mobile recharge purchases and cell phone type; mobility features such as radius of gyration and number of places visited; and social features such as interactions per contact and entropy of contacts (roughly, the randomness of a person's communication patterns).

After extracting the relevant features, existing studies link the feature to a related socioeconomic outcome. In estimating commuting patterns in urban centers in Haiti, Zagatti et al. (2018) identified home and workplace locations for mobile phone users using call frequency data and an unsupervised learning algorithm. They found that the most relevant features for identifying 'meaningful' locations were the number of distinct days in which a mobile phone user connected to a particular cell tower, and the time between the first and last of these *call-days*. In another application centered around mobility, Lu et al. (2016) used mobile data to model migration in Bangladesh after the landfall of Tropical Cyclone Mahasen. They defined a *move* as a user who generated at least two CDRs more than 10 kilometers apart during a six-hour period on the day of the landfall. Following the 2010 earthquake in Haiti, Bengtsson et al. (2011) used call location data to track the movement of mobile phone users and found that the net outflow of people from the capital was equal to 20 percent of the city population.

Importantly, several studies have predicted the sex of mobile phone users, in addition to other relevant socioeconomic variables. Al-Zuabi et al. (2019) use CDRs to predict the sex and age of mobile phone users in Syria. They found that the entropy of call duration and the number of contacts were among the most informative CDR-derived features for predicting sex, with their gradient boosting model achieving an accuracy of 85.6 percent. Jahani et al. (2017) used a lasso model to identify the most discriminative features between sexes. They found that, in an unnamed European country, the delay in response time to a text was the most distinguishing feature, as men exhibited shorter response times. In contrast, in

the South Asian country studied, the percentage of calls made by the user during weekend nights was the most discriminative feature. Their model achieved prediction accuracies of about 74 percent in both countries. In contrast to the complex features derived by Jahani et al. (2017), Felbo et al. (2017) achieved similar accuracy in predicting gender using only raw CDR data and a convolutional neural network, without incorporating mobility information.

Beyond accurately predicting gender, some studies have used CDR data to explore and shed light on dynamics relevant to women's empowerment and gender gaps. Blumenstock (2012) examined internal migration in Rwanda using a CDR dataset alongside phone survey data that collected the mobile phone numbers, demographic information, and movement history of 901 people. He found only insignificant differences in migration levels based on sex, which contrasts with most literature that suggest men are more likely to migrate. Similarly, Dalberg Data Insights (2019) used phone survey data along with CDRs and found that women have a lower radius of gyration than men, suggesting reduced mobility. They also found that while more women use mobile money than men, men tend to engage more frequently in individual mobile-money services like deposits, withdrawals and total transactions.

Of particular relevance to our paper are several studies using CDR data to predict wealth and poverty. Based on the hypothesis that users who make few large purchases of airtime credit are wealthier than those who make many smaller purchases, Gutierrez et al. (2013) built social networks from CDRs in Côte d'Ivoire and found that users tend to be in social networks with others with similar purchase patterns. Steele et al. (2017) combined CDR data and satellite data to build highly accurate poverty maps of Bangladesh, achieving fine spatial granularity down to the neighborhood level, and in some cases, households and individuals.

Unlike this spatial approach (despite the inclusion of some individual-level prediction), other studies solely predict individual-level outcomes from individual-level phone data, broadening the potential applications of the results. For example, Blumenstock et al. (2015) merged CDRs of 1.5 million subscribers with phone survey data from 856 people in Rwanda, using machine-learning algorithms to predict the wealth of individuals, achieving a cross-validated correlation coefficient of 0.68. Similarly, Aiken et al. (2022) matched CDR data with survey data in Togo to refine the targeting of assistance during the COVID-19 pandemic. Compared to other methods based on geography, occupation, and CDR data alone, their approach that combined machine-learning and CDR data significantly reduced the errors of exclusion in targeting the assistance.

3. Data

We use three sources of data: a phone-based survey with MTN subscribers in Ghana, de-identified⁶ digital trace data of MTN subscribers who participated in the phone survey, and geospatial data linked to the subscriber data via cell tower locations. At the time of our analysis, MTN—the largest mobile network operator in Ghana—served nearly 24 million customers, representing 60 percent of voice subscriptions and 70 percent of mobile data subscriptions in the country (Ghana National Communications Authority 2022).

⁶ De-identified data refers to data that has been stripped of personally identifiable information so that individuals cannot be directly identified from the data. This process typically involves removing or masking names, addresses, national identification numbers, and other unique identifiers (via www.edglossary.org).

3.1 Phone survey

In February 2022, we conducted a phone survey of randomly selected MTN subscribers in Ghana. From a database of over 27 million subscribers, phone numbers were randomly selected using a sampling probability proportional to the population distribution in the MTN database, stratified by rural–urban status across each of 16 administrative regions. The goal was to complete at least 8,000 interviews, with a minimum of 40 percent women interviewed. This sample size was determined to be the minimal threshold necessary to achieve a representative sample suitable for training the machine-learning model, based on previous work (GSMA 2018). All ethical protocols were adhered to during data collection, management and analysis, and was approved by the Institutional Review Board of the Alliance of Bioversity International and CIAT.

The survey was carefully designed to minimize the burden on respondents and the likelihood of refusal to participate. It included 18 questions about respondents' work and employment over the past seven days, along with an additional four questions about their mobile phone usage (see [Appendix 1](#)). We aligned the questions with ILO (2013) standards concerning statistics of work, employment,⁷ and labor underutilization. We also informed the questions by consulting with labor statisticians from the Ghana Statistical Service (GSS). Additional questions were included to verify the respondent's age, sex and whether they were the primary user of the current SIM card (i.e., the person who uses the phone more than anyone else). Enumerators had instructions to attempt (up to three times) to arrange a time to speak with the primary or main user.

Although the legal working age in Ghana is 15 years, we restricted the survey to respondents aged 18 years or older due to ethical concerns related to interviewing minors. Thus, in this study, we define the working-age population as people aged 18 years or older.

A team of 36 experienced enumerators conducted the survey, operating out of a call center set up by GSS for this survey, in observance with prevailing COVID-19 protocols. Based on GSS's prior experience with phone surveys in Ghana (where respondents often questioned the legitimacy of enumerators, partly due to an increase in fraud during the COVID-19 pandemic (BBC News Africa 2020)), GSS recommended sending a text message to prospective respondents. This message informed them of their potential selection for an upcoming phone survey conducted by GSS, and included a phone number for an official GSS contact that respondents could call to verify the survey's legitimacy. Approximately one week prior to the start of fieldwork, GSS sent the text messages to prospective respondents. Informal communications with GSS indicate that only a small proportion of respondents called the GSS contact number to inquire about the survey.

Out of the total 13,053 respondents contacted, we successfully completed interviews with 8,162 respondents, or 62.5 percent—a response rate that is much higher than what is typically achieved for phone surveys such as ours (i.e., not conducted in connection with a previous in-person survey). We attribute the high response rate partly to the short duration of the survey (the median interview time was 6 minutes, 55 seconds), but mostly to the overall effectiveness of GSS's text message strategy in increasing respondents' confidence in the legitimacy and importance of the survey.

Following the removal of 96 respondents during data cleaning because of various inconsistencies, the final analytic sample includes 8,066 respondents, consisting of 3,468 women (43 percent) and 4,598 men (57 percent). These data enable us to measure individual-level labor-market participation, specifically employment, unemployment and underemployment.

[Table 1](#) provides detailed descriptions of these three indicators as defined in our study and their official definitions by the ILO (2013). The reference period for all indicators is the

7 According to ILO, *employment* is one form of work among several others, which include own-use production, unpaid trainee work, volunteer work, and other work activities (ILO 2013).

past seven days. ‘Employment’ refers to any activity to produce goods or provide services, whether for pay, profit or own consumption, including temporary absences from such work. ‘Unemployment’ is defined as not being employed but actively searching for a job and being available to work. ‘Underemployment’ includes both unemployed and employed people who worked fewer than 40 hours in the past seven days and who were available to work additional hours.

Table 1. Summary of labor-market indicators to be predicted

Indicator	Definition used in our study	ILO (2013) definition
Employed	Respondent <u>worked with or without remuneration for at least one hour anytime during the past seven days</u> in household farming, livestock, or fishing activities, in a nonfarm business, or as an employee or apprentice; or if respondent did not work in the last seven days they <u>currently have a job, business or family farm to which they plan to return</u> .	Employed persons are defined as all those of working age who, during the last seven days, were engaged in any activity to produce goods or provide services for pay or profit. They comprise: a. employed persons “ <i>at work</i> ”, i.e. who worked in a job for at least one hour b. employed persons <i>not at work</i> due to temporary absence from a job, or to working-time arrangements
Unemployed	Respondent <u>did not work with or without remuneration anytime during the past seven days</u> in household farming, livestock or fishing activities, in a nonfarm business, or as an employee or apprentice; and they <u>did nothing to find a paid job or start a business and would not have started working had a job or business opportunity been available</u> .	Unemployed persons are defined as all those of working age who, during the last seven days, did not engage in any activity to produce goods or provide services for pay or profit, but did carry out activities to seek employment and were available to take up employment given a job opportunity.
Underemployed	Respondent <u>worked with or without remuneration for less than 40 hours during the past seven days</u> in household farming, livestock or fishing activities, in a nonfarm business, or as an employee or apprentice; and they <u>were available and willing to work more hours</u> .	Underemployed persons are defined as all those of working age who, in the last seven days, were engaged for at least one hour (but less than 40 hours) in any activity to produce goods or provide services for pay or profit, and who, during the last seven days, wanted to work additional hours and were available to work additional hours given an opportunity.

Table 2 provides descriptive statistics for the labor-market indicators and mobile phone usage variables. The data show that both male and female respondents are, on average, around 35 years old. Women are less likely to be employed and more likely to be unemployed than men. Approximately one-fourth of respondents report being underemployed, with no significant gender difference. Excluding unpaid work and domestic activities (which were not measured in our survey) women worked, on average, nearly eight hours less than men during the past seven days. These findings are consistent with estimates from the seventh round of the Ghana Living Standards Survey (GLSS7), a nationally representative household survey conducted between October 2016 and October 2017.⁸

Regarding phone usage, women are slightly more likely to need permission to use the phone, though this is rare for both genders. Both women and men commonly carry their phones outside the home; men are slightly more likely to always carry theirs. Most people use phones for personal matters, but about one-third of respondents also report using their

⁸ Specifically, national-level estimates calculated using GLSS7 data indicate that, among women aged 18 years and older, 70.7 percent were employed, 3.7 percent were unemployed and 14.6 percent were underemployed compared to 76.3, 4.8 and 14.9 percent respectively for men aged 18 years and older. The difference compared to our estimates are, thus, relatively small and can be attributed to differences in sampling and in the exact phrasing of survey questions.

phones for business purposes, with men being slightly more likely to do so than women. Additionally, men are more likely to use multiple phones or SIM cards.

Table 2. Comparison of labor-market indicators and other variables for women and men

Variable	Men		Women		Difference
	Mean	SD	Mean	SD	
Employed	0.84	0.37	0.72	0.45	0.12 ***
Unemployed	0.07	0.25	0.10	0.29	-0.03 ***
Underemployed	0.25	0.44	0.24	0.43	0.02
Age	35.20	12.30	34.70	13.10	0.50
Must ask permission to use phone	0.07	0.26	0.08	0.28	-0.01 *
Phone is ...always carried	0.87	0.34	0.83	0.38	0.04 ***
... sometimes carried	0.12	0.32	0.16	0.36	-0.04 ***
... rarely carried	0.01	0.09	0.01	0.10	-0.00
... never carried	0.01	0.07	0.01	0.09	-0.00
Phone is ...never used for work	0.62	0.49	0.71	0.45	-0.10 ***
... sometimes used for work	0.32	0.47	0.25	0.43	0.07 ***
... exclusively used for work	0.06	0.25	0.04	0.19	0.03 ***
Does not use another mobile phone	0.49	0.50	0.64	0.48	-0.15 ***
Observations	4,598		3,468		

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

3.2 Digital trace data

The digital trace data for the sampled MTN subscribers cover a six-month period from November 2021 to May 2022. Previous research suggests that a minimum reference period of three months is sufficient to adequately predict subscribers' usage behavior (GSMA 2018). However, due to potential disruptions in usage caused by the COVID-19 pandemic during the reference period, we opted for a longer six-month period for this study. Nonetheless, as discussed below, it remains an open question whether this extended reference period was sufficient to adequately establish subscribers' usage patterns amid the rapidly changing conditions during that time.

[Table 3](#) summarizes the types of digital trace data used in the analysis. This data includes behavioral information regarding calling, texting, and data usage (collectively referred to as CDRs); mobile money transactions; and airtime purchases. Each activity instance is accompanied by the time stamp and location metadata, based on the cell tower that processed the activity. Notably, the sex of the subscriber is unknown and must be predicted alongside our other outcomes of interest.

Prior to analysis, the raw digital trace data, which consist of multiple entries per subscriber, are transformed into numerical *features*, where each feature has a single value per subscriber that can then be processed by the machine-learning models. In total, we computed 204 features. Examples of these features include the number of voice calls or text messages sent or received during the reference period, as detailed in [Table 3. Appendix 2](#) provides a complete list of features and their descriptions. The selection of these features was informed by prior research that identified them as promising candidates for predicting sex and socioeconomic status (GSMA 2018).

Table 3. Summary of digital trace data and features

Types of data		Trace data available	Examples of features calculated based on trace data
Call-detail records	Voice call	Phone number for calls made and received, duration of calls, cell tower, time stamp	Number of voice calls, total number of incoming/outgoing voice calls, average duration of voice calls on weekdays/weekends
	Text message	Phone number for messages sent and received, cell tower, time stamp	Number of incoming text messages to which the user responded within 60 minutes, average delay between the time of receiving and responding to a text message
	Data usage	Amount of data traffic, time stamp, cell tower	Number of data transactions, volume of data used
Mobile money		Phone number for payments sent and received, transaction amount, transaction fee/commission, cell tower, timestamp	Number of days with mobile money transactions, value sent/received via P2P transaction, minimum/maximum account balance
Airtime purchases (top-ups)		Purchase amount, payment channel, cell tower, time stamp	Number of instances SIM card was topped-up, average value of top-ups
Physical mobility		Derived from above	Number of different cell towers visited, radius of gyration (average distance traveled)
Social networking		Derived from above	Number of unique contacts, number of unique contacts for all incoming/outgoing voice/text activity

Note: The database includes the longitude and latitude for each identified cell tower. See [Appendix 2](#) for a full description of features.

3.3 Geospatial data

Geospatial data were obtained from public databases and processed for all of Ghana at spatial resolutions ranging from 500 m to 5 km, depending on the source. These data include environmental and physical metrics likely to be associated with labor-market behavior, such as satellite-detectable nighttime lighting, climatic conditions, distance to cities and ports, and land cover. [Appendix 3](#) provides a detailed summary of the geospatial features used in the analysis. These geospatial data were linked to the digital trace data based on the location of subscribers' "home" cell tower, the one they most frequently connected to during the reference period.

In total, the training data contain records from all 2,849 cell towers operated by MTN in Ghana during the reference period. [Figure 1](#) illustrates the approximate locations of these cell towers, which are predominantly situated in the more populous southern region of Ghana—particularly the metropolitan areas around the two largest cities, Kumasi and Accra.

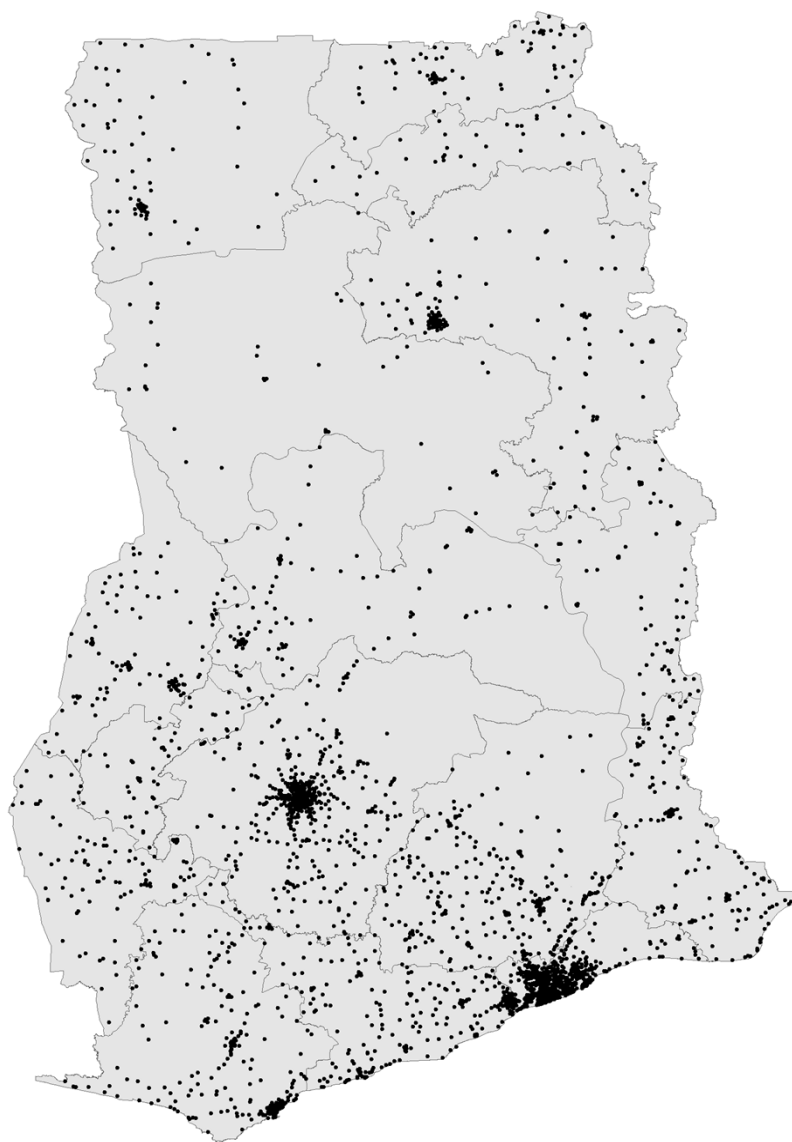


Figure 1. Approximate cell tower locations for phone survey sample (n = 2,849)

4. Methodology

We use the extracted digital trace and geospatial features, along with phone survey data, to train supervised machine-learning models to jointly predict mobile phone subscribers' sex and their employment, unemployment, and underemployment status based on their historical mobile phone usage. As noted, the joint prediction approach distinguishes our study from previous research, which typically focuses on predicting individual socioeconomic characteristics independently of the subscriber's sex and vice versa. In this section, we describe in detail the three competing prediction models used in our analysis, then explain the cross-validation exercises used to assess model performance.

4.1 Prediction models

For each outcome, we generate predictions using three models: Logistic Regression (LR), Bayesian Additive Regression Tree (BART), and Shared Forest (SF). The BART and SF are supervised machine-learning models designed to handle the joint prediction problem in distinct ways. The LR serves as a benchmark for comparison against these machine-learning models.

4.1.1 Baseline model

First, we consider a baseline framework to predict employment indicators conditional on gender. Let g_i and y_i represent the binary gender and employment indicator of interest of the i^{th} surveyed individual respectively. Further, let x_i represent the corresponding vector of features calculated from the CDR. We fit two independent LRs, specifying:

$$\begin{aligned} g_i &\sim \text{Bernoulli}(\rho_i) \\ \log\left(\frac{\rho_i}{1-\rho_i}\right) &= x_i^T \beta \end{aligned} \quad \text{and} \quad \begin{aligned} y_i &\sim \text{Bernoulli}(\phi_i) \\ \log\left(\frac{\phi_i}{1-\phi_i}\right) &= x_i^T \gamma \end{aligned} \quad (1)$$

where the vectors β and γ are estimated using maximum likelihood estimation.

The above framework can be used to obtain out-of-sample sex-disaggregated predictions of the employment indicator based solely on CDR data. The first LR yields out-of-sample predictions of gender, $\hat{g}_i = I(\hat{\rho}_i \geq 0.5)$,⁹ where $\hat{\rho}_i = (1 + e^{-x_i^T \hat{\beta}})^{-1}$, for any input vector of features, x . Then, we can obtain out-of-sample predictions of employment indicators by averaging values of $\hat{\phi}_i = (1 + e^{-x_i^T \hat{\gamma}})^{-1}$ separately for $\hat{g}_i = 0$ and $\hat{g}_i = 1$ cases. For example, male indicators can be estimated as:

$$E(\phi \mid g = 0) = \frac{1}{|\{i: \hat{g}_i = 0\}|} \sum_{i: \hat{g}_i = 0} \hat{\phi}_i. \quad (2)$$

The framework described has several limitations. First, these models depend on the researcher to choose the variables included in the covariate vector x . Relatedly, they also do not automatically incorporate interactions between covariates, requiring the researcher to identify any heterogeneous relationships and explicitly include them. Furthermore, the models assume that the relationships between the parameter of interest and the covariates are monotonic and follow a logistic curve, which may not always be appropriate.

4.1.2 BART

We propose using an ensemble of classification trees as the basis for the modeling approach because trees do not suffer the aforementioned limitations. In particular, tree structures are

9 $I()$ represents the indicator function where $I(A)$ evaluates to 1 if A is true, and 0 otherwise.

desirable because they perform automatic variable selection by choosing to incorporate, or split by, only those variables that sufficiently reduce error variability. They also handle interactions and model highly nonlinear relationships through sequential splitting on variables, which is particularly useful when dealing with many features.

Figure 2 illustrates these concepts in the context of a classification tree predicting the probability corresponding to a binary response, y , using two covariates labeled X_1 and X_2 . The left panel illustrates the sequential binary splits of the tree, while the right panel illustrates how the tree partitions the covariate space to produce probability predictions. For an example, if $X_1 = 0.25$ and $X_2 = 0.50$, we can see that this observation in the tree structure would trickle down from the root node (top of the tree) down to the leaf node labeled 5. Note that X_2 has a discontinuous and positive effect on the predicted probability—that $y = 1$, but only for small values of X_1 . This heterogeneous effect is an example of how trees naturally incorporate interactions and nonlinear relationships.

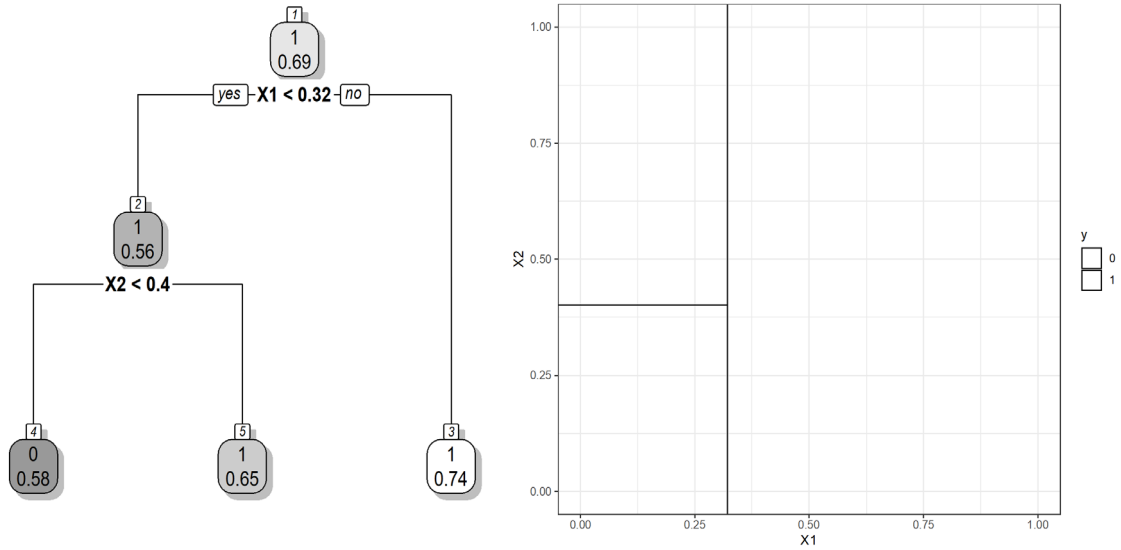


Figure 2. Illustration of a decision tree predicting a binary response, y

The main drawback of single trees is that they tend to overfit data, resulting in poor out-of-sample predictions (James et al. 2013). We employ a sum-of-trees model, known as a BART, developed by Chipman, George and McCulloch (2007) and Chipman, George and McCulloch (2010). BART enjoys the same benefits of trees but tends to be less likely to overfit by aggregating the predictions of many trees and limiting the influence of any one tree through regularization priors.

We use a set of two independent BARTs to model gender and employment indicators in a way that parallels the logistic regressions described in the previous section. Letting g_i again represent gender, we still model g_i with a Bernoulli distribution. However, instead of linking the logit of ρ_i to a linear function of covariates, we set:

$$\Phi^{-1}(\rho_i) = \sum_{t=1}^T f_t(x_i | T_t, M_t) \quad (3)$$

where Φ^{-1} represents the probit link function and $f_t(\cdot)$ represents the t^{th} tree function.

$f_t(\cdot)$ inputs a vector of features and outputs a prediction on the probit scale. The output of $f_t(\cdot)$ also depends on T_t , the collection of binary decision rules within the tree, and M_t , the prediction parameters. Supposing Figure 2 is an illustration of the first tree in a BART model, T_1 would contain the binary decision rules $X_1 < 0.32$ and $X_2 < 0.40$ while M_1 would contain the vector (0.58, 0.65, .74). This tree would explain a small portion of the variability in the response, (T_2, M_2) would explain a portion of the remaining variability, and this pattern would continue sequentially.

The estimation of the sex-conditional employment indicators follows a similar approach to that used with the logistic regression model described in the previous section. The binary employment indicator adopts the same specification as above, though with its own tree structures T_t and prediction parameters M_t . The final predictions can be made using equation (2).

The BART model offers greater flexibility compared to the logistic regressions specified in equation (1) in terms of modeling covariates. However, both models are univariate and thus ignore potentially useful information about the joint relationship between sex and employment outcomes. This is particularly relevant given the substantial evidence demonstrating the highly gendered nature of employment. Both models treat the prediction of mobile phone subscribers' sex and their employment/unemployment/underemployment status as independent classification problems.

4.1.3 Shared forest

Information on what predicts sex could potentially be used to additionally inform what predicts employment outcomes, and vice versa. This kind of information-borrowing is leveraged in the bivariate SF model proposed by Linero, Sinha and Lipsitz (2020). The SF model incorporates information on the joint relationship between sex and employment outcomes by modeling the two responses simultaneously using a shared tree structure. If the two responses are driven by or associated with similar features, then the SF model is expected to generate more precise predictions compared to the BART model.

In an SF model, the two responses are modeled simultaneously and share tree structures, T_t , dictating the sequential binary splits. If the two responses are driven by or associated with similar features, the model is more precise in estimating the underlying tree structures. Importantly, the two responses maintain independent prediction parameters, M^p and M^y . The original framework of Linero, Sinha and Lipsitz (2020) was built to incorporate at least one nonbinary response variable. We modify it to simultaneously model two binary responses—gender and employment indicators—resulting in a SF model that can be written as:

$$\begin{aligned} g_i &\sim \text{Bernoulli}(\rho_i) \\ y_i &\sim \text{Bernoulli}(\gamma_i) \\ \Phi^{-1}(\rho_i) &= \sum_{t=1}^T f_t(x_i | T_t, M_t^p) \\ \Phi^{-1}(\gamma_i) &= \sum_{t=1}^T f_t(x_i | T_t, M_t^y). \end{aligned} \tag{4}$$

Although this model allows for information-sharing across gender and employment indicators, it represents a simplified version of the model specified in the previous section. Thus, predictive gains are likely only if that simplification (a shared tree structure) is warranted.

4.2 Estimation of machine-learning models

The BART and SF models are both machine-learning models estimated using Bayesian techniques, which require the specification of priors. A prior represents our probabilistic beliefs about a parameter before seeing or analyzing any data. It controls how the model behaves and, in a prediction setting, helps to prevent excessive complexity or “overfitting.” For our ensemble models, we place priors on T (the tree structures) and M (terminal node parameters). $P(T)$ is formed by assigning prior probabilities to any given node of depth d being nonterminal though the equation $\alpha(1 + d)^{-\beta}$ for $\alpha \in (0,1), \beta \in (0,\infty)$. A common and default choice for these hyperparameters is to set $\alpha = 0.95$ and $\beta = 2$, which places relatively high weight on trees with two or three terminal nodes. The parameters within M are assigned conjugate normal priors; we direct the reader to Chipman, George and McCulloch (2010) for details on the default hyperparameters in the normal distributions. We set the number of trees to 100 and sample 10,000 draws from the posterior distribution, discarding the first 5,000 as burn-in. For both the BART and SF models, we opted to use default hyperparameter values because they are known to perform relatively well, and tuning them would have been computationally prohibitive given the multiple replications in our study.

4.3 Cross-validation

We assess model performance using cross-validation statistics, which involve computing performance metrics across multiple samples and averaging them to obtain a more reliable measure, thereby reducing bias from particular train/test splits. To do this, we begin by creating 30 random train/test splits with 70 percent of the data put into training and 30 percent put into corresponding test sets. The LR, BART and SF models are independently estimated on each of the 30 training sets. Predictions in the form of probabilities are then obtained for each model on the 30 corresponding test sets.

Using the predicted probability across the 30 test sets, along with the corresponding true values of gender and employment, we calculate sensitivity and specificity values across a range of classification thresholds. Sensitivity measures the model's accuracy of predicting positive outcomes, while specificity measures the accuracy of it predicting negative outcomes (i.e., the probability that observed positive or negative cases are correctly classified). We then plot the sensitivity and specificity pairs for each model to construct receiver operator characteristic (ROC) curves and calculate the area under the curve (AUC).

AUC is a measure of a model's overall ability to correctly distinguish between positive and negative results. Figure 3 illustrates two ROC curves and their corresponding AUCs as shaded areas; we plot pairs of true-positive rates (sensitivity) vs false-positive rates (1- specificity) for two hypothetical models of differing quality. Labels indicate the 'optimal' specificity and sensitivity based on maximizing the sum of the two metrics. Figure 3 suggests that Model 1 is a better predictor because it is able to achieve higher sensitivities and specificities, which is reflected in a higher AUC value. Given a randomly chosen positive and negative case, AUC can also be interpreted as the probability that the model will rank the positive response higher than the negative one (i.e., that it would be able to correctly classify the two). It ranges from 0 (incorrectly predicts 100 percent of cases) to 1 (correctly predicts 100 percent of cases), where 0.5 is equivalent to random guessing. AUC is favored over the more commonly used accuracy in cases of imbalanced data (such as unemployment). We report AUC along with the optimized sensitivity and specificity pair.

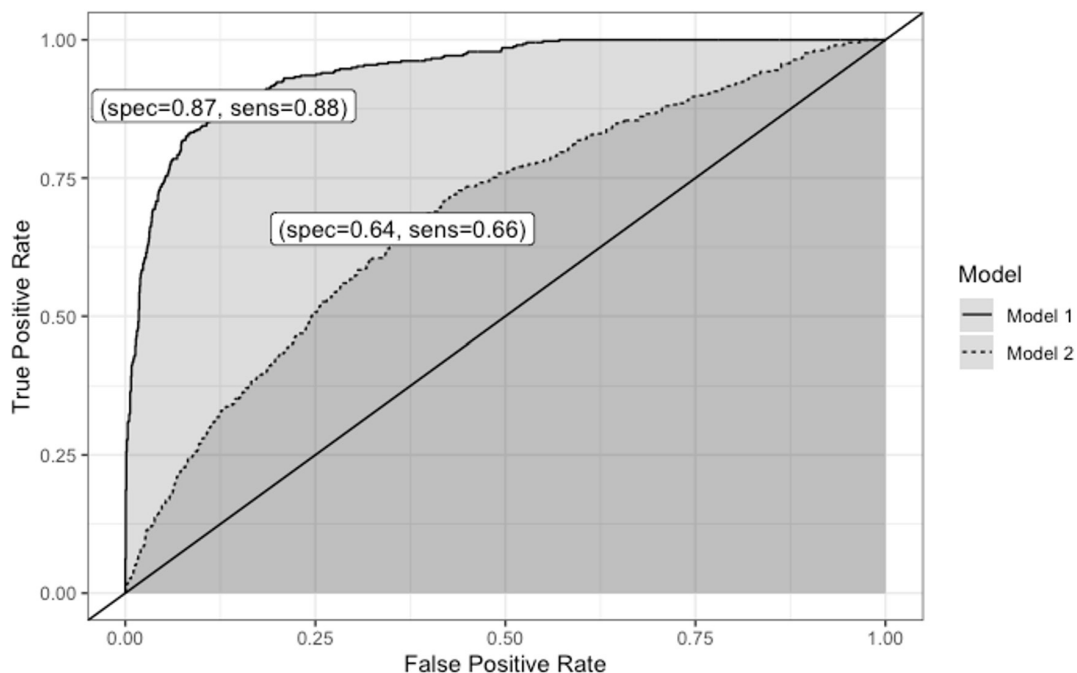


Figure 3. Illustration of two hypothetical ROC curves

5. Results

5.1 Cross-validation of predictions

Table 4 presents cross-validation statistics assessing the ability of each model to correctly predict the sex of the mobile phone subscriber. We find that all three models predict mobile phone subscribers' sex at roughly the same level, achieving results consistent with previous findings (e.g., Slavchevska et al. 2021). Specifically, all three models correctly classify mobile phone subscribers as women in 70–71 percent of the cases and as male subscribers in 64–66 percent of the cases. As an overall performance metric, the AUC ranges from 0.72 to 0.73 across all models.

Table 4. Cross-validation statistics showing each model's ability to predict mobile phone subscribers' sex

Model	Sensitivity	Specificity	AUC
LR	0.696	0.656	0.728
BART	0.695	0.642	0.722
SF	0.707	0.636	0.724

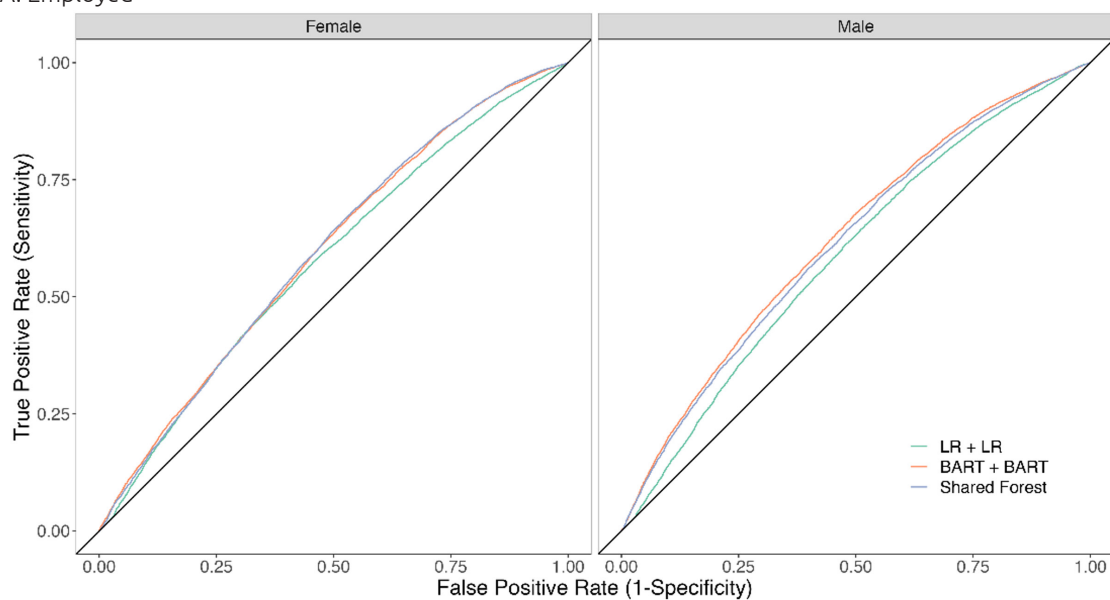
Note: LR = logistic regression; BART = Bayesian additive regression tree; SF = Shared forest; AUC = area under the curve.
Source: authors' calculations.

Figure 4 presents ROC curves showing each model's ability to jointly predict mobile phone subscribers' sex and their employment, unemployment, and underemployment status. Corresponding cross-validation statistics are shown in Table 5. Relative to the models' ability to predict mobile phone subscribers' sex alone, we observe a steep decline in model performance. The models would be expected to correctly classify the employment, underemployment, or unemployment status of female subscribers in 52–59 percent of cases and male subscribers in 52–58 percent of cases, with the AUC ranging from 0.54 to 0.63 across models and outcomes. Introducing the joint prediction problem thus leads to a 10–19 percentage point decrease in sensitivity, a 3–15 percentage point decrease in specificity for both women and men, and a decrease in AUC ranging from 0.10 to 0.19.

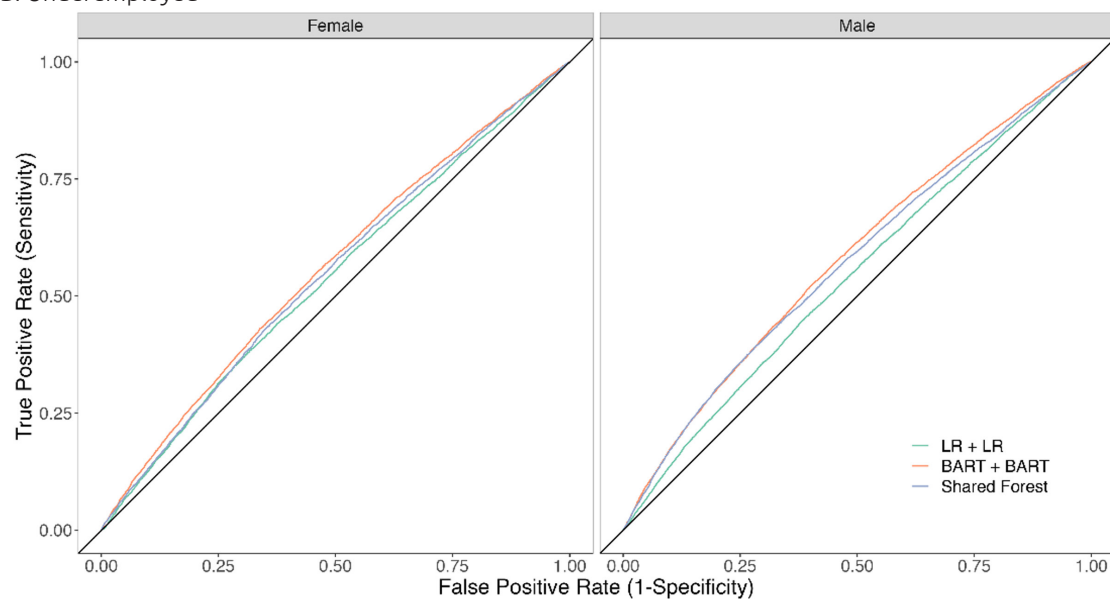
Although the three models achieved similar levels of performance in predicting mobile phone subscribers' sex alone, this parity disappears with the introduction of the joint prediction problem. The BART and SF models consistently outperform LR, in terms of AUC, by 0.02–0.04, regardless of outcome or sex.

When comparing the BART and SF models, we observe little difference in model performance across outcomes, but notable variation in model performance by sex. The BART and SF models perform equally well when jointly predicting unemployment status for either sex, with an AUC of approximately 0.59. However, model performance diverges for employment and underemployment status. For women, AUC ranges from 0.55 to 0.60, while for men, it ranges from 0.57 to 0.63. These values are demonstrably lower than the 70.4 percent accuracy reported by Sundsøy et al. (2017) in predicting individual unemployment, albeit for all users—not disaggregated by sex.

A. Employed



B. Underemployed



C. Unemployed

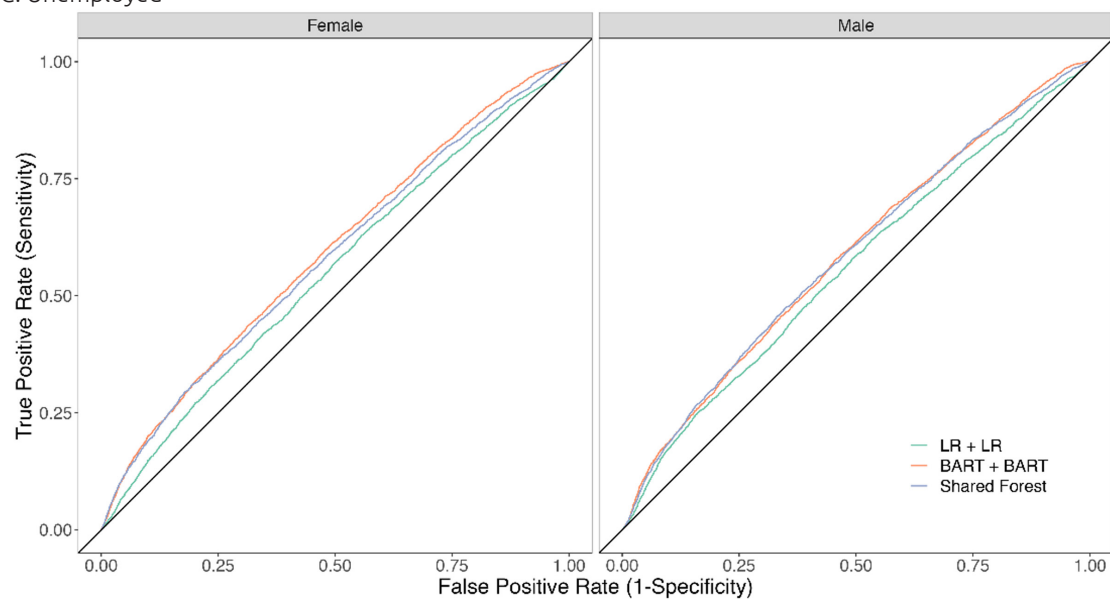


Figure 4. ROC curves showing each model's ability to predict female and male mobile phone subscribers' employment, unemployment, and underemployment status. Source: authors' calculations.

Table 5. Cross-validation statistics showing each model's ability to predict employment, unemployment, and underemployment status by the sex of mobile phone subscribers. Source: authors' calculations.

Women	Outcome	Model	Sensitivity	Specificity	AUC
	Employed	LR + LR	0.523	0.559	0.582
		BART + BART	0.527	0.569	0.596
		SF	0.527	0.570	0.597
	Underemployed	LR + LR	0.529	0.509	0.538
		BART + BART	0.547	0.515	0.561
		SF	0.536	0.511	0.547
	Unemployed	LR + LR	0.549	0.507	0.556
		BART + BART	0.585	0.509	0.594
		SF	0.573	0.508	0.581

Men	Outcome	Model	Sensitivity	Specificity	AUC
	Employed	LR + LR	0.515	0.579	0.594
		BART + BART	0.520	0.605	0.625
		SF	0.518	0.597	0.615
	Underemployed	LR + LR	0.532	0.511	0.543
		BART + BART	0.562	0.521	0.583
		SF	0.554	0.518	0.572
	Unemployed	LR + LR	0.556	0.506	0.561
		BART + BART	0.582	0.506	0.588
		SF	0.581	0.506	0.587

5.2 Prediction mapping

To illustrate how this methodology could be used to address gender data gaps in individual-level labor-market statistics, we used the BART model to generate out-of-sample, sex-disaggregated predictions for the unemployment status of the 19.5 million MTN subscribers who did not participate in the phone survey.

Although the SF model slightly outperformed BART, we opted for BART due to computational limitations. Combined with cell tower locations in the digital trace data, these predictions enable us to study the geographic distribution of unemployment at an exceptionally fine level of spatial detail. Traditional survey-based data on labor-market outcomes for Ghana are typically representative only at the regional level; digital trace data allow us to predict the unemployment status for women and men across 4,256 microregions which approximate mobile phone coverage zones. These microregions were generated using Voronoi tessellation and cell tower locations. [Figure 5](#) highlights the difference in spatial resolution, contrasting the Voronoi polygons (black borders) with Ghana's 16 administrative regions (green borders).

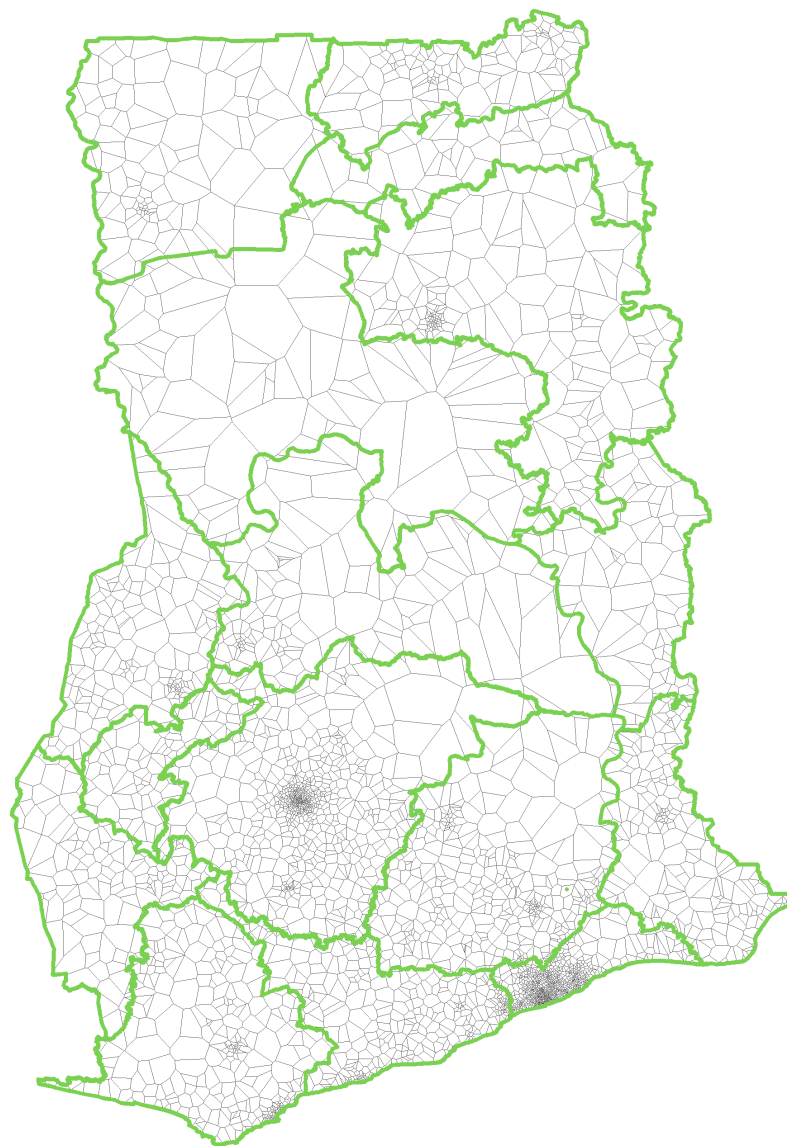


Figure 5. Spatial structure of Voronoi polygons (in black) based on cell tower locations compared to regional borders (in green) in Ghana. Source: authors' calculations.

We populated the map using out-of-sample, sex-disaggregated predictions from the BART model. Each polygon was assigned a value corresponding to the unemployment rate among female and male subscribers who “reside” within that coverage zone. The resulting geographical distributions display the predicted unemployment rate among female subscribers in [Figure 6](#) and the gender differences in predicted unemployment rates between female and male subscribers (calculated as the female unemployment rate minus the male unemployment rate) in [Figure 7](#). Both figures include an inset focusing on the area around the capital city, Accra.

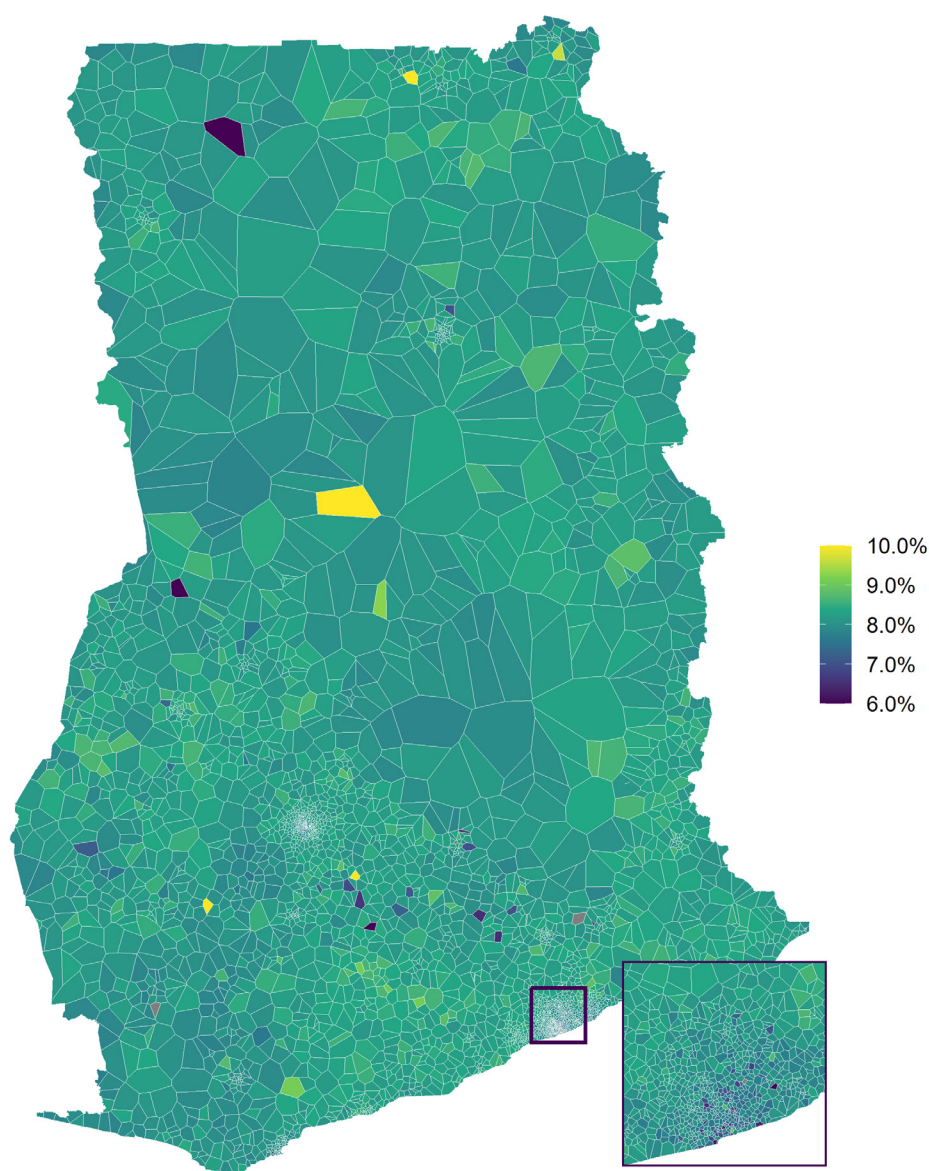


Figure 6. National-level prediction map for the unemployment rate among female MTN subscribers based on digital trace and geospatial data. Source: authors' calculations.

The prediction map for the unemployment rates among female subscribers (Figure 6) reveals a high degree of geographical homogeneity. Except for a few coverage zones, with a few exceptions, the predicted unemployment rate among female subscribers remains relatively stable across coverage zones, consistently ranging between 7.7 percent (5th percentile) and 8.5 percent (95th percentile). The overall unemployment rate for female subscribers is 8.2 percent, which is slightly lower than the 9.5 percent rate observed in the phone survey sample.

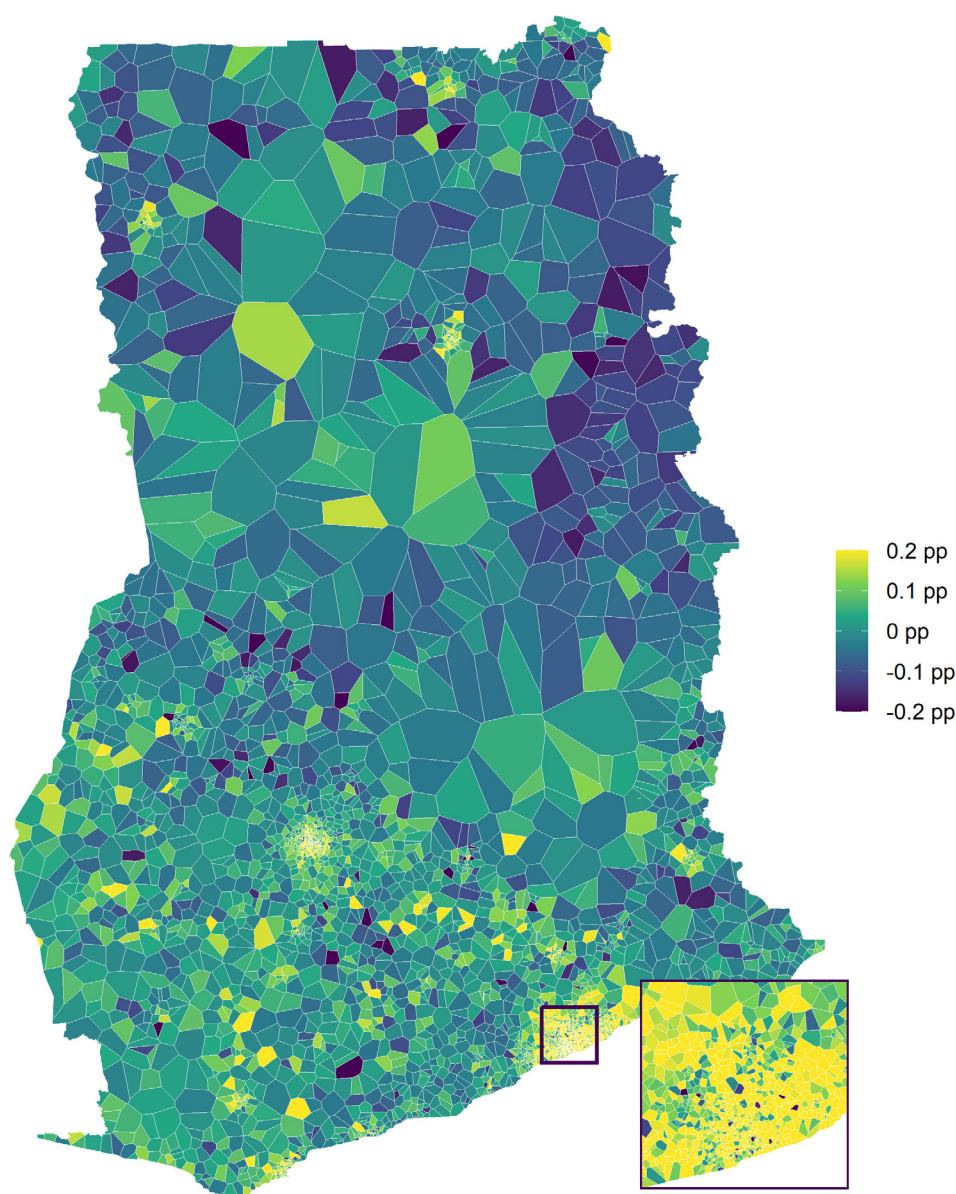


Figure 7. National-level prediction map for gender differences in the unemployment rates among female and male MTN subscribers, based on digital trace and geospatial data.

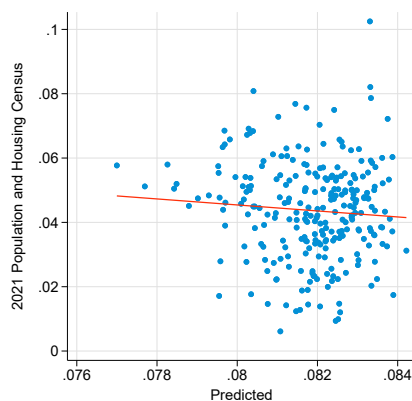
Note: Gender difference = female unemployment rate – male unemployment rate; pp = percentage points. Source: authors' calculations.

In comparison, the prediction map for the gender difference in unemployment rates (Figure 7) shows a greater degree of spatial variation. The magnitudes of the predicted gender differences are small, ranging from 0.08 (5th percentile) to 1.1 percentage points (95th percentile). The overall average gender gap is 0.06 percentage points, considerably less than the 2.6 percentage points gender gap observed in the phone survey sample. Nevertheless, there are pockets of coverage zones within the map where gender differences appear more pronounced. For example, urban centers like Accra in the southeast and Kumasi in central Ghana contain a relatively higher concentration of coverage zones where the mean unemployment rate for female subscribers exceeds that of males. Conversely in the northeast, particularly within the Northern Region, there are more coverage zones where the mean unemployment rate among female subscribers is less than that of males. This spatial variation, which is largely absent from the prediction map that displays the mean unemployment rates among female subscribers, suggests that these predictions may indeed reflect meaningful insights about gendered labor-market dynamics in Ghana.

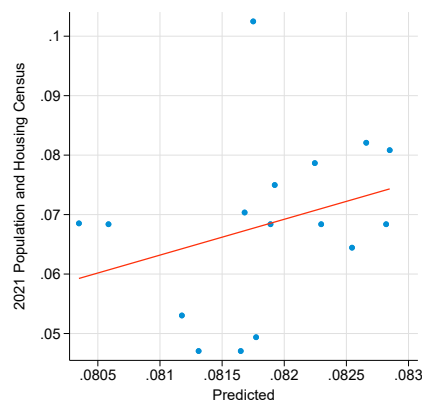
5.3 External validation

To determine whether the observed variation in the predicted unemployment rates accurately reflects gendered labor dynamics in Ghana, it is essential to validate our predictions using existing data or official statistics. However, this validation is complicated by the lack of publicly available national-level unemployment data in Ghana that matches the spatial resolution of our predictions. The most comparable dataset available is Ghana's 2021 Population and Housing Census,¹⁰ from which only a 10 percent random subsample is publicly accessible via the GSS. While these data do not allow for a direct comparison at the *coverage zone*-level used in our prediction maps, they do permit comparison at the district and regional levels.

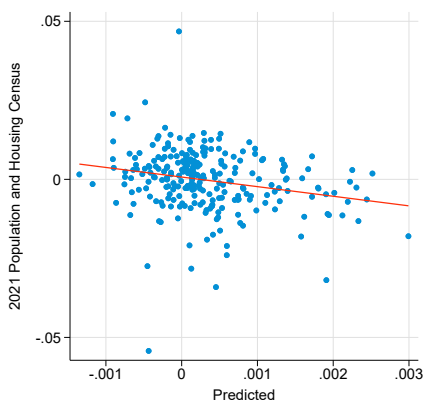
Scatter plots showing these comparisons are presented in Figure 8. Panels A and B compare the predicted unemployment rates among female subscribers across 261 districts or 16 administrative regions, plus the mean unemployment rate among women and men based on the 2021 census data. Panels C and D compare the predicted mean difference in the unemployment rate among female and male subscribers within each district or region, plus the mean difference in the unemployment rate among women and men based on the 2021 census data.



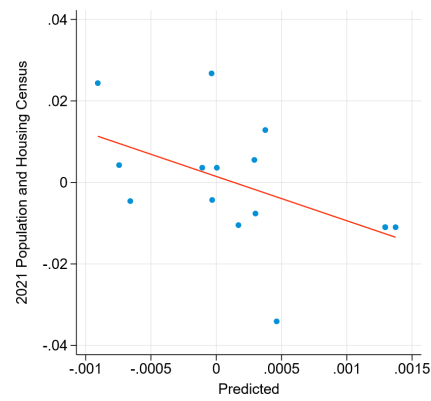
Panel A. Women's unemployment rate by district ($\rho = -0.08$)



Panel B. Women's unemployment rate by region ($\rho = 0.30$)



Panel C. Gender difference in unemployment rate by district ($\rho = -0.23$)



Panel D. Gender difference in unemployment rate by region ($\rho = -0.50$)

Figure 8. External validation of predicted unemployment rates among women and gender differences in unemployment rates using data from Ghana's 2021 Population and Housing Census. Source: authors' calculations.

Note: linear prediction in red; gender difference = women's unemployment rate – men's unemployment rate.

¹⁰ <https://www2.statsghana.gov.gh/nada/index.php/catalog/110>

Overall, the model predictions do not hold up well in these external comparisons. At the regional level, the unemployment rates for women show a positive correlation ($\rho = 0.30$), but this relationship does not extend to the district level where the correlation is negative ($\rho = -0.08$). Furthermore, the mean gender differences in unemployment rates are negatively correlated at both the district ($\rho = -0.23$) and regional ($\rho = -0.50$) levels. These findings suggest that the model is largely unsuccessful in accurately predicting gendered labor-market dynamics in Ghana.

6. Discussion and conclusions

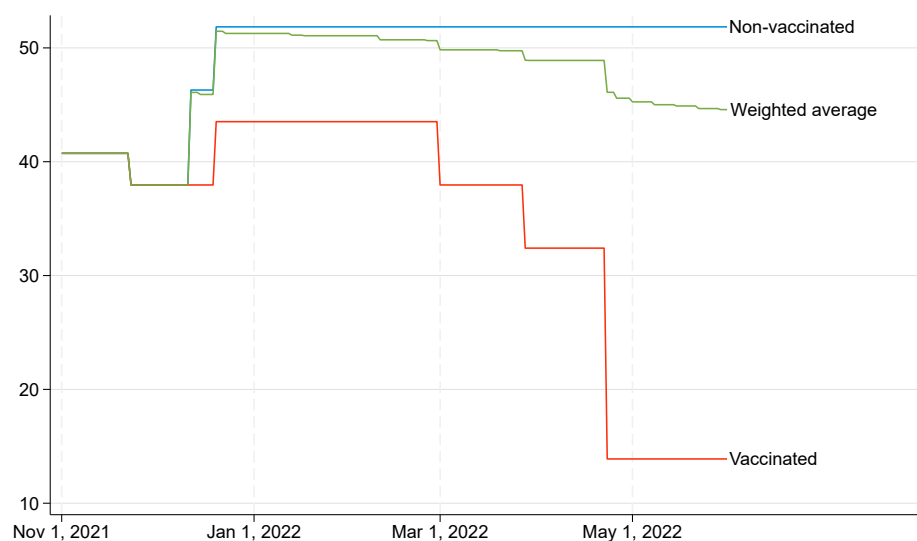
Since our study is the first to apply machine-learning models to predict gendered labor-market outcomes, it is difficult to assess the performance of our prediction models. In cross-validation, our models predict labor-market outcomes only slightly better than random guessing, which is equivalent to an AUC of 0.5. The BART model also underperforms in external validation, showing minimal correlation with comparable estimates based on census data. Both results are less favorable than those reported in related literature. For example, Blumenstock, Cadamuro and On (2015) achieved AUC values ranging from 0.72 to 0.81 and demonstrated a strong correlation with survey-based composite wealth indices when predicting the poverty status of mobile phone users in Rwanda.

Our study differs from others in this field for two main reasons. First, we tackle a complex joint-prediction problem in predicting gender-specific outcomes simultaneously, which presents methodological challenges not encountered in studies focusing on individual socioeconomic characteristics independently. Second, our focus on labor-market indicators (employment, unemployment and underemployment) is relatively novel and less explored compared to other outcomes like poverty status or migration patterns. The underperformance of our methodology may stem from several factors. Our understanding of how mobile phone usage varies (or not) among employed, unemployed or underemployed individuals (both women and men) is relatively limited. Minimal variation in mobile phone usage across these categories could also hinder the effectiveness of training the machine-learning models and thus negatively impact model performance. Additionally, the limited variation across observed outcomes, such as the small gender differences observed in the labor-market indicators (see [Table 1](#)), may have similar effects. Simply put, labor-markets indicators may not be as well suited to prediction using digital trace data compared to other outcomes studied in the literature.

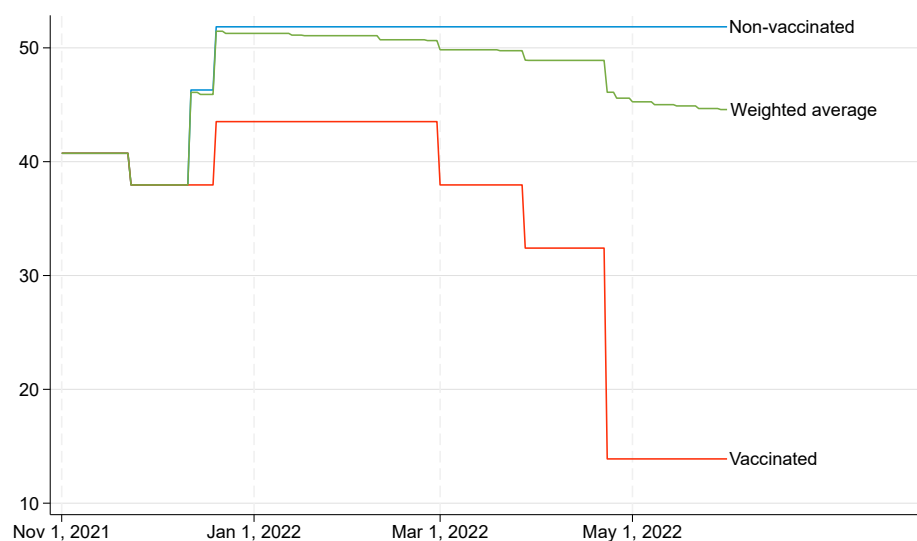
The data used in our study may also be abnormally noisy. This is because the reference period for both the labor and digital trace data used in our study overlaps with a particularly tumultuous phase of the COVID-19 pandemic in Ghana. This period was marked by significant disruptions, which likely introduced considerable variability and noise into the data, potentially impacting model performance. These disruptions are summarized in [Figure 9](#). The data sources include the Oxford Covid-19 Government Response Tracker project's Stringency Index (Hale et al. 2021) and Google COVID-19 Community Mobility Reports (Google, LLC 2022).

In November 2021, near the beginning of our reference period, COVID-19 vaccines had only recently become widely available in Ghana, and in December the government reported its first case of the Omicron COVID-19 variant. The Stringency Index ([Figure 9](#), Panel A) reflects an abrupt increase in restrictions just before January 2022, while the Google COVID-19 Community Mobility Reports ([Figure 9](#), Panel B) show sharp decreases in movement to grocery stores and pharmacies, retail and recreation locations, transit stations and workplaces around the same time. COVID-19 restrictions remained strict until early March

2022, midway through the reference period for our study. Afterward, restrictions gradually eased, particularly for vaccinated individuals. Mobility rebounded much quicker and remained relatively stable over the remainder of the reference period, with only a few sharp but short-lived decreases in movement to workplaces in late April and early May.



Panel A. COVID-19 Stringency Index



Panel B. Google COVID-19 Community Mobility Reports

Figure 9. COVID-19 policy response in Ghana from November 1, 2021, to May 31, 2022. Source: Hale et al. (2021) and Google, LLC (2022)

Note: The Stringency Index aggregates information from nine policy response metrics, based on publicly available data and consistently coded to capture variation in degree of response.¹¹ It ranges from 0 to 100, with higher values reflecting stricter responses. Since policies often differ by vaccination status, the Stringency Index is calculated for three categories: vaccinated people, nonvaccinated people, and a weighted average based on the share of vaccinated people. Google's COVID-19 Community Mobility Reports use information on visits and length of stay to show daily mobility trends across different six categories of places relative to a baseline day, the median daily value between January 3 and February 6, 2020.¹²

11 The nine metrics used to calculate the Stringency Index are: school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements, and international travel controls.

12 <https://www.google.com/covid19/mobility/>

The drastically changing conditions experienced by respondents in Ghana during the reference period are very likely to have affected their mobile phone usage. Previous studies analyzing different types of shocks have shown similar effects in different contexts. For example, Blumenstock et al. (2016) observed increased airtime transfers following the 2008 Lake Kivu earthquake in Rwanda, and Toole et al. (2015) linked changes in call volume to the closure of a large manufacturing plant. The most compelling evidence of these effects comes from Katsumata et al. (2022), who studied smartphone usage in Japan during an earlier phase of the COVID-19 pandemic. By comparing data from 2019 to 2020, they found that the frequency of smartphone usage during daytime hours on weekdays increased following the declaration of a state of emergency, with a more significant increase for women compared to men. This demonstrates that not only did the change in COVID-19 restrictions affect overall smartphone usage, but it also had differential effects on women's and men's smartphone usage patterns.

In summary, it is reasonable to expect that gender differences in mobile phone behavior were fluctuating over the course of the reference period. As a result, features in the digital trace data that might have been strong predictors of gender and employment at other times might not have performed as effectively in our sample. While certainly this is not the only factor, it may explain why our study achieved lower accuracy levels compared to other research studies. With these considerations in mind, we believe the methodology still holds promise for filling gender data gaps in individual-level labor-market statistics. Despite the sub-optimal accuracy, our results illustrate that the methodology offers a credible approach for producing national statistics at a finer degree of spatial granularity than traditional household surveys. This capability is valuable for filling data gaps.

Future work could benefit from a broader and more diverse set of training data to mitigate disruption-related issues similar to the COVID-19 issues we encountered. Additionally, refining the methodology used for feature engineering and selection is crucial. Computational limitations led us to engineer and select features based on insights from prior similar work (GSMA 2018), but alternative methods—such as those implemented by Blumenstock, Cadamuro and On (2015), which involve creating thousands of metrics per user—may offer more robust and expansive results, albeit with higher computational demands. Both refinements could significantly improve performance while maintaining the same spatial granularity in the predicted statistics. Moreover, we recommend expanding the range of labor-market indicators in future research—such as disaggregating between sectors of employment (e.g., agricultural versus non-agricultural)—which may reveal greater variation in mobile phone usage among working women and men, and increase the quality of predictions.

References

- Aiken, Emily, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E. Blumenstock. 2022. "Machine Learning and Phone Data Can Improve Targeting of Humanitarian Aid." *Nature* 603 (7903): 864–70. <https://doi.org/10.1038/s41586-022-04484-9>.
- Al-Zuabi, Ibrahim Mousa, Assef Jafar, and Kadan Aljoumaa. 2019. "Predicting Customer's Gender and Age Depending on Mobile Phone Data." *Journal of Big Data* 6 (1): 18. <https://doi.org/10.1186/s40537-019-0180-9>.
- Ambel, Alemayehu, Kevin McGee, and Asmelash Tsegay. 2021. "Reducing Bias in Phone Survey Samples: Effectiveness of Reweighting Techniques Using Face-to-Face Surveys as Frames in Four African Countries (English)." World Bank, Policy Research Working Paper, WPS 9676. Washington, D.C.: World Bank Group.
- BBC News Africa. 2020. "Corona Quacks: Exposing Fake Coronavirus Cures in Ghan." <https://www.youtube.com/watch?v=qX0jbLxFQ90>.
- Bengtsson, Linus, Xin Lu, Anna Thorson, Richard Garfield, and Johan von Schreeb. 2011. "Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti." *PLoS Medicine* 8 (8): 1–9. <https://doi.org/10.1371/journal.pmed.1001083>.
- Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350 (6264): 1073–76. <https://doi.org/10.1126/science.aac4420>.
- Blumenstock, Joshua E. 2012. "Inferring Patterns of Internal Migration from Mobile Phone Call Records: Evidence from Rwanda." *Information Technology for Development* 18 (2): 107–25. <https://doi.org/10.1080/02681102.2011.643209>.
- Blumenstock, Joshua E., Nathan Eagle, and Marcel Fafchamps. 2016. "Airtime Transfers and Mobile Communications: Evidence in the Aftermath of Natural Disasters." *Journal of Development Economics* 120: 157–181. <https://doi.org/10.1016/j.jdevco.2016.01.003>.
- Brubaker, Joshua, Talip Kilic, and Philip Wollburg. 2021. "Representativeness of Individual-Level Data in COVID-19 Phone Surveys: Findings from Sub-Saharan Africa." *PLoS ONE* 16 (11): e0258877. <https://doi.org/10.1371/journal.pone.0258877>.
- Buvinić, Mayra, Rebecca Furst-Nichols, and Gayatri Koolwal. 2014. "Mapping Gender Data Gaps." New York: Data2x. <https://data2x.org/resource-center/mapping-gender-data-gaps/>.
- Chipman, Hugh A., George, Edward I., and Robert E. McCulloch. 2010. "BART: Bayesian Additive Regression Trees." *The Annals of Applied Statistics* 4 (1): 266–98. <http://www.jstor.org/stable/27801587>.
- Chipman, Hugh A., George, Edward I., and Robert E. McCulloch. 2007. "Bayesian ensemble learning." *Neural Information Processing Systems* 19: 265–272.
- Dalberg Data Insights. 2019. "Differences in Mobile Money and Phone Usage between Men and Women in Uganda: Case Study." Dalberg Data Insights. https://data2x.org/wp-content/uploads/2019/12/MobileMoneyinUganda_DDI.pdf.
- Data2x and Kate Grantham. 2020. "Mapping Gender Data Gaps: An SDG Era Update." New York: Data2x. https://data2x.org/wp-content/uploads/2020/03/MappingGenderDataGaps_ExecSummary.pdf.
- Desiere, Sam, and Valentina Costa. 2019. "Employment Data in Household Surveys: Taking Stock, Looking Ahead." World Bank, Policy Research Working Paper 8882. Washington, DC: World Bank. <https://doi.org/10.1596/1813-9450-8882>.
- Eagle, Nathan, Michael Macy, and Rob Claxton. 2010. "Network Diversity and Economic Development." *Science* 328 (5981): 1029–31. <https://doi.org/10.1126/science.1186605>.
- Felbo, Bjarke, Pål Sundsøy, Alex Sandy Pentland, Sune Lehmann, and Yves Alexandre de Montjoye. 2017. "Modeling the Temporal Nature of Human Behavior for Demographics Prediction." In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Skopje, Macedonia, September 2017. https://doi.org/10.1007/978-3-319-71273-4_12.

- Ghana National Communications Authority. 2022. "Industry Information - Telecom Subscriptions for January 2022." National Communications Authority. <https://nca.org.gh/wp-content/uploads/2022/06/Industry-Information-January-2022-1.pdf>.
- Google, LLC. 2022. "COVID-19 Community Mobility Reports." Google. <https://www.google.com/covid19/mobility>.
- Gourlay, Sydney, Talip Kilic, Antonio Martuscelli, Philip Wollburg, and Alberto Zezza. 2021. "Viewpoint: High-Frequency Phone Surveys on COVID-19: Good Practices, Open Questions." *Food Policy* 105 (December): 102153. <https://doi.org/10.1016/j.foodpol.2021.102153>.
- GSMA. 2018. "The Gender Analysis & Identification Toolkit: Estimating Subscriber Gender Using Machine Learning." London: GSM Association. <https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/wp-content/uploads/2018/08/GSMA-Gender-Analysis-and-Identification-Report-August-2018.pdf>.
- GSMA. 2022. "The Mobile Gender Gap Report 2022." London: GSMA. <https://www.gsma.com/r/gender-gap-2022/>.
- Gutierrez, Thoralf, Gautier Krings, and Vincent D. Blondel. 2013. "Evaluating Socio-Economic State of a Country Analyzing Airtime Credit and Mobile Phone Datasets." arXiv. <https://doi.org/10.48550/arXiv.1309.4496>.
- Hale, Thomas, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, et al. 2021. "A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker)." *Nature Human Behaviour* 5 (4): 529–38. <https://doi.org/10.1038/s41562-021-01079-8>.
- ILO. 2013. "Resolution Concerning Statistics of Work, Employment and Labour Underutilization." In 19th International Conference of Labour Statisticians, Geneva, Switzerland, October 2013. <https://www.ilo.org/meetings-and-events/19th-international-conference-labour-statisticians>.
- Jahani, Eaman, Pål Sundsøy, Johannes Bjelland, Linus Bengtsson, Alex Sandy Pentland, and Yves Alexandre de Montjoye. 2017. "Improving Official Statistics in Emerging Markets Using Machine Learning and Mobile Phone Data." *EPJ Data Science* 6 (1). <https://doi.org/10.1140/epjds/s13688-017-0099-3>.
- James, G., Witten, D., Hastie, T., and R. Tibshirani. 2013. *An introduction to statistical learning*. 303–332. New York: Springer.
- Katsumata, Sotaro, Takeyasu Ichikohji, Satoshi Nakano, Shinichi Yamaguchi, and Fumihiko Ikuine. 2022. "Changes in the Use of Mobile Devices during the Crisis: Immediate Response to the COVID-19 Pandemic." *Computers in Human Behavior Reports* 5 (March): 100168. <https://doi.org/10.1016/j.chbr.2022.100168>.
- Lavelle-Hill, Rosa, John Harvey, Gavin Smith, Anjali Mazumder, Madeleine Ellis, Kelefa Mwantimwa, and James Goulding. 2022. "Using Mobile Money Data and Call Detail Records to Explore the Risks of Urban Migration in Tanzania." *EPJ Data Science* 11 (1): 28. <https://doi.org/10.1140/epjds/s13688-022-00340-y>.
- Linero, A. R., Sinha, D., and S. R. Lipsitz. 2020. "Semiparametric mixed-scale models using shared Bayesian forests." *Biometrics* 76 (1): 131–144. <https://doi.org/10.1111/biom.13107>.
- Lopes, Claudia Abreu, and Savita Bailur. 2018. "Gender Equality and Big Data." New York: UN Global Pulse; UN Women. https://doi.org/10.1007/978-94-007-0753-5_1129.
- Lu, Xin, David J. Wrathall, Pål Roe Sundsøy, Md Nadiruzzaman, Erik Wetter, Asif Iqbal, Taimur Qureshi, et al. 2016. "Unveiling Hidden Migration and Mobility Patterns in Climate Stressed Regions: A Longitudinal Study of Six Million Anonymous Mobile Phone Users in Bangladesh." *Global Environmental Change* 38 (May): 1–7. <https://doi.org/10.1016/j.gloenvcha.2016.02.002>.
- Pestre, Gabriel, Emmanuel Letouzé, and Emilio Zagheni. 2020. "The ABCDE of Big Data: Assessing Biases in Call-Detail Records for Development Estimates." *The World Bank Economic Review* 34 (Supplement 1): S89–97. <https://doi.org/10.1093/wber/lhz039>.
- Slavchevska, Vanya, Marcelo Tyszler, Dharani Dhar Burra, Greg Seymour, Denys Sementsov, Astrid Van Lierde, and Brian King. 2021. "Can Call Detail Records Provide Insights into Women's Empowerment? A Case Study from Uganda." CGIAR Platform for Big Data in Agriculture; CGIAR Research Program on Policies, Institutions, and Markets (PIM). <https://hdl.handle.net/10568/111023>.

- Steele, Jessica E., Pål Roe Sundsøy, Carla Pezzulo, Victor A. Alegana, Tomas J. Bird, Joshua Blumenstock, Johannes Bjelland, et al. 2017. "Mapping Poverty Using Mobile Phone and Satellite Data." *Journal of the Royal Society Interface* 14 (127). <https://doi.org/10.1098/rsif.2016.0690>.
- Sundsøy, Pål, Johannes Bjelland, Bjørn-Atle Reme, Eaman Jahani, Erik Wetter, and Linus Bengtsson. 2017. "Towards Real-Time Prediction of Unemployment and Profession." In *Social Informatics. SocInfo 2017. Lecture Notes in Computer Science*, Vol. 10540, edited by G. Ciampaglia, A. Mashhadi, and T. Yasseri, 14–23. Springer, Cham. https://doi.org/10.1007/978-3-319-67256-4_2.
- Toole, J. L., Lin, Y. R., Muehlegger, E., Shoag, D., González, M. C., and D. Lazer. 2015. "Tracking employment shocks using mobile phone data." *Journal of the Royal Society Interface* 12 (107).
- Zagatti, Guilherme Augusto, Miguel Gonzalez, Paolo Avner, Nancy Lozano-Gracia, Christopher J. Brooks, Maximilian Albert, Jonathan Gray, et al. 2018. "A Trip to Work: Estimation of Origin and Destination of Commuting Patterns in the Main Metropolitan Regions of Haiti Using CDR." *Development Engineering* 3 (January): 133–65. <https://doi.org/10.1016/J.DEVENG.2018.03.002>.

Appendix 1. Phone survey questionnaire

Module B: Respondent information

B_START. Start of module B date and time stamp:	DD	MM	YYYY	HH	MM	SS
---	----	----	------	----	----	----

Interviewer: Record a new attempt every time you call a number.

B01	Did anyone answer the phone?	1 = Yes 2 = No >> end of survey
B02	INFORMED CONSENT STATEMENT	1 = Yes 2 = No >> end of survey
B03	Are you the main user of this handset and/or SIM card? By main user, I mean do you use it more than anybody else. [Interviewer: if the respondent does not understand the word SIM card, then please read the definition] SIM cards are inserted into a mobile phone handset or other device to enable customers to access a mobile operator's network. The SIM card may have been installed in the handset by the supplier of the phone and it can come in different sizes. Every SIM has a unique telephone number.]	1 = Yes 2 = No >> end of survey
B04	How should I greet you?	1 = Madam/Miss 2 = Sir
B05	What is your age?	(years)
IF AGE LESS THAN 18 YEARS >>> END OF SURVEY		
B06	Do you have to ask permission to use the handset/SIM?	1 = Yes 2 = No
B07	How often do you carry this handset on your person when outside of the home?	1 = Always 2 = Sometimes 3 = Rarely 4 = Never
B08	Do you regularly conduct work on the phone such as selling products, ordering inputs, connecting with customers, etc.?	1 = Yes, exclusively used for work 2 = No, sometimes used for work 3 = No, never used for work
B09	Is this the only handset and/or SIM you use?	1 = Yes 2 = No, I have multiple handsets/SIMs

B_END. End of module B date and time stamp:	DD	MM	YYYY	HH	MM	SS
---	----	----	------	----	----	----

Module C: Work and employment

C_START. Start of module C date and time stamp:	DD	MM	YYYY	HH	MM	SS
---	----	----	------	----	----	----

EMPLOYMENT		
C01	In the last 7 days, that is from [DAY, DATE] to [DAY, DATE], did you work with or without remuneration for at least one hour in household farming, livestock, or fishing activities; in a non-farm business; or as an employee or apprentice?	1 = Yes 2 = No >>> Skip to C16
C02	In what work activity did you spend most of your time in the last 7 days? DO NOT READ OPTIONS	1 = Working on household farming, livestock, or fishing activities 2 = Running a non-farm business of any size for yourself or household >>> Skip to C04 3 = Helping without pay in a non-farm business >>> Skip to C04 4 = Working as an employee for a wage, salary, commission, or remuneration of any kind >>> Skip to C04 5 = Paid apprenticeship >>> Skip to C04 6 = Unpaid apprenticeship >>> Skip to C04
C03	Thinking about the products obtained from your farming, livestock or fishing activity, are they intended mainly for sale or mainly for family use or consumption?	1 = Mainly for sale 2 = Mainly for family use or consumption
C04	In the last 7 days, on how many days did you work on your primary work activity?	(days)
C05	How many hours per day, on average, did you spend working on your primary work activity?	(hrs)
C06	Thinking about the total hours that you worked in the last 7 days on your primary work activity, is this less than usual, the same as usual, or more than usual?	1 = Less than usual 2 = Same as usual 3 = More than usual
[CAPI Instructions: Proceed to C07 only if C04*C05 = less than 40 hours. Otherwise, skip to C09]		
UNDEREMPLOYMENT		
C07	Thinking about the total hours that you worked in the last 7 days on your primary work activity, would you have been available to work more hours?	1 = Yes 2 = No
C08	Thinking about the total hours that you worked in the last 7 days on your primary work activity, would you have been willing to work more hours?	1 = Yes 2 = No
INFORMALITY		
C09	Do you have a written or verbal contract or agreement or letter of appointment for your primary work activity?	1 = Written agreement 2 = Verbal agreement 3 = No
C10	In your primary work activity, are you provided health insurance coverage?	1 = Yes 2 = No
C11	In your primary work activity, are you entitled to paid leave of any kind (e.g., vacation, holiday, sick, or maternity/paternity leave)?	1 = Yes 2 = No
C12	In your primary work activity, are you entitled to social security benefits?	1 = Yes, all of these 2 = Yes, some of these 2 = No, none of these

COMMUTING		
C13	In the morning, how many minutes does it typically take to travel between your home and the place where you work for your primary work activity?	(min)
C14	In the evening, how many minutes does it typically take to travel between your home and the place where you work for your primary work activity?	(min)
C15	What mode of transportation do you use the most when travelling to and from the place where you work for your primary work activity? DO NOT READ OPTIONS	1 = Walking 2 = Bicycle 3 = Private automobile 4 = Motorbike/tricycle 5 = Taxi/hurry car 6 = Tro tro 7 = Bus 8 = Train >>> End of module
[CAPI Instructions: Proceed to C16 only if C01 = No. Otherwise, proceed to next module]		
TEMPORARY ABSENCE		
C16	Even though you did not work in the last 7 days, do you currently have a job, business, or family farm that you plan to return to?	1 = Yes >>> End of module 2 = No
UNEMPLOYMENT		
C17	During the last 7 days, did you do anything to find a paid job or start a business?	1 = Yes 2 = No
C18	During the last 7 days, if a job or business opportunity had been available, could you have started working last week?	1 = Yes 2 = No

Appendix 2. Mobile phone usage features used in the analysis

Types of data		Features analyzed
Call-detail records	Voice call	Total number of voice/text/data activities Total number of voice calls Total number of voice calls on weekdays/weekends Total number of voice calls during the morning/afternoon/evening on weekdays/weekends Total number of incoming/outgoing voice calls Total number of incoming/outgoing voice calls on weekdays/weekends Total number of incoming/outgoing voice calls during the morning/afternoon/evening on weekdays/weekends Total duration of voice calls Total duration of voice calls on weekdays/weekends Total duration of voice calls during the morning/afternoon/evening on weekdays/weekends Total duration of incoming/outgoing voice calls Total duration of incoming/outgoing voice calls on weekdays/weekends Total duration of incoming/outgoing voice calls during the morning/afternoon/evening on weekdays/weekends Average duration of voice calls Average duration of voice calls on weekdays/weekends Average duration of voice calls during the morning/afternoon/evening of weekdays/weekends Average duration of incoming/outgoing voice calls Average duration of incoming/outgoing voice calls on weekdays/weekends Average duration of incoming/outgoing voice calls during the morning/afternoon/evening of weekdays/weekends Proportion of outgoing voice calls (total number of outgoing voice calls divided by total number of voice calls) Proportion of outgoing voice call duration (total duration of outgoing voice calls divided by total duration of voice calls)
	Text message	Total number of text messages Total number of text messages on weekdays/weekends Total number of text messages during the morning/afternoon/evening on weekdays/weekends Total number of incoming/outgoing text messages Total number of incoming/outgoing text messages on weekdays/weekends Total number of incoming/outgoing text messages during the morning/afternoon/evening on weekdays/weekends Total number of incoming text messages to which the user responded within 60 minutes Total number of incoming text messages to which the user responded within 60 minute / total number of incoming text messages (ratio) Average delay between the time of receiving and responding to a text message
	Data usage	Total number of data transactions Total number of data transactions on weekdays/weekends Total number of data transactions during the morning/afternoon/evening on weekdays/weekends Average duration of data transactions Average duration of data transactions on weekdays/weekends Average duration of data transactions during the morning/afternoon/evening on weekdays/weekends Total volume of data used Total volume of data used on weekdays/weekends Total volume of data used during the morning/afternoon/evening on weekdays/weekends

Types of data	Features analyzed
Mobile money	Binary indicator of whether the subscriber has a mobile account Binary indicator of whether subscriber received more mobile money than sent via peer-to-peer (P2P) transaction Total number of days with mobile money transactions Total number of days user sent/received mobile money Total number of unique types of mobile money transactions sent/received Total number of unique accounts with which user has transacted Total number of mobile money transactions sent/received Total number of airtime transactions Total number of merchant payment transactions Total number of deposits/withdrawals Total number of P2P transactions sent/received Total number of P2P transactions sent/received during weekdays/weekends Total number of P2P transactions sent/received during each day of the week Total value of mobile money sent/received Total value of mobile money fees Total value of airtime transactions Total value of merchant payment transactions Total value deposited/withdrawn Total value sent/received via P2P transaction Total value sent/received via P2P transaction during weekdays/weekends Total value sent/received via P2P transaction during each day of the week Average value of mobile money transactions sent/received Average value of mobile money fees Average value of airtime transactions Average value of merchant payment transactions Average value deposited/withdrawn Average value sent/received via P2P transaction Average account balance Minimum/maximum account balance Minimum/maximum value deposited/withdrawn
Data purchases (top-ups)	Total number of days during which SIM card was topped-up Total number of instances SIM card was topped-up Total number of top-ups made via electronic voucher drive (EVD) Total number of top-ups made via mobile money Total value of top-ups Average value of top-ups Average value of top-ups per active day Average value of top-ups made via EVD Average value of top-ups made via mobile money
Physical mobility	Urban/rural classification of home area (based on the location of the cell tower that routes the majority of the user's activity) Total number of different cell towers visited Total number of different level-3 administrative divisions visited Total number of different level-2 administrative divisions visited Total number of different level-1 administrative divisions visited Radius of gyration, i.e., the average distance traveled; calculated following (Rodríguez-Crespo et al. 2012) Average number of unique towers visited per active day Average number of different level-3 administrative divisions visited per active day Average number of different level-2 administrative divisions visited per active day Average number of different level-1 administrative divisions visited per active day Average radius of gyration per active day
Social networking	Total number of unique contacts Total number of unique contacts for incoming/outgoing voice calls Total number of unique contacts for incoming/outgoing text messages Total number of unique contacts for all incoming/outgoing voice/text activity Average number of unique contacts per active day

Appendix 3. Summary information for remote sensing–GIS datasets used for covariate selection

Type	Description	Source	Spatial resolution	Temporal resolution
Climate	Rainfall	Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) version 2.0 Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., Romero, B. E., Husak, G. J., Michaelsen, J. C., and A. Verdin. 2014. <i>A quasi-global precipitation time series for drought monitoring</i> . US Geological Survey Data Series 832, 4. http://pubs.usgs.gov/ds/832/ .	3 arc minute (~5 km at the equator)	Monthly; November 2021 – May 2022
Climate	Average temperature; maximum temperature; minimum temperature	Berkeley Earth Gridded Daily Land Temperature Rohde, R. A., and Z. Hausfather. 2020. "The Berkeley Earth Land/Ocean Temperature Record." <i>Earth System Science Data</i> 12 (4): 3469–3479. https://doi.org/10.5194/essd-12-3469-2020 .	60 arc minute (~100 km at the equator)	Monthly; November 2021 – May 2022
Human settlement	Nighttime lights	Visible and Infrared Imaging Suite (VIIRS) Monthly Cloud-free Day Night Band (DNB) Composite Elvidge, C. D., K. E. Baugh, M. Zhizhin, and F.-C. Hsu. 2013. "Why VIIRS data are superior to DMSP for mapping nighttime lights." <i>Proceedings of Asia-Pacific Advanced Network</i> 35: 62.	15 arc second (~500 m at the equator)	Monthly; Nov 2021 – May 2022
Human settlement	Travel time to cities; ¹³ travel time to ports ¹⁴	Version 3 (2019-05-15) from https://figshare.com/articles/dataset/Travel_time_to_cities_and_ports_in_the_year_2015/7638134/3 Nelson, A., Weiss, D. J., van Etten, J. et al. 2019. "A suite of global accessibility indicators." <i>Sci Data</i> 6: 266. https://doi.org/10.1038/s41597-019-0265-5 .	30 arc second (~1 km at the equator)	2015

¹³ Travel times are divided into nine classes based on city population: 5,000–10,000; 10,000–20,000; 20,000–50,000; 50,000–100,000; 100,000–200,000; 200,000–500,000; 500,000–1,000,000; 1,000,000–5,000,000; 5,000,000–50,000,000.

¹⁴ Travel times are divided into five classes based on port size: large, medium, small, very small, and any.

Type	Description	Source	Spatial resolution	Temporal resolution
Land cover	Proportion of landcover class (trees, grassland, shrubs, cropland, built, bare, snow, water, wetland, mangroves, moss)	ESA Worldcover 2020 v100 Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, et al. 2021. <i>European Space Agency (ESA) WorldCover 10 m 2020 v100</i> . https://doi.org/10.5281/zenodo.5571936 .	30 arc second (~1 km at the equator)	2020
Livestock	Estimated count of cattle, sheep, goats, horses, pigs, chickens, and ducks	Gridded Livestock of the World (GLW v4) Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T. P., Vanwambeke, S. O., Wint, G. R. W., and T. P. Robinson. 2018. "Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010." <i>Scientific Data</i> 5: 1–11. https://doi.org/10.1038/sdata.2018.227 .	5 arc minute (~10 km at the equator)	2015
Population	Estimated population count by age-sex group ¹⁵	WorldPop Open Population Repository (https://wopr.worldpop.org/) Leasure, D. R., Darin, E., and A. J. Tatem. 2021. "Bayesian gridded population estimates for Ghana 2019 (GHA v2.0)." <i>WorldPop, University of Southampton</i> . https://doi.org/10.5258/SOTON/WP00705 .	3 arc second (~90 m at the equator)	2019

¹⁵ The data are divided into 36 categories for males and female, respectively (72 categories in total). For young children, these include categories for males and females under 1-year and aged 1- to 4-years. Between ages 5 and 80, the data are grouped in five-year bins for males and females (e.g., 5- to 9-years, 10- to 14-years, etc.). Males and female ages 80 years and above are represented in a single category, respectively.



Generating Evidence and New Directions for Equitable Results (GENDER) is CGIAR's impact platform designed to put equality and inclusion at the forefront of global agricultural research for development. The Platform is transforming the way gender research is done, both within and beyond CGIAR, to kick-start a process of genuine change toward greater gender equality and better lives for smallholder farmers everywhere.

gender.cgiar.org



CGIAR is a global research partnership for a food-secure future dedicated to reducing poverty, enhancing food and nutrition security, and improving natural resources.

cgiar.org