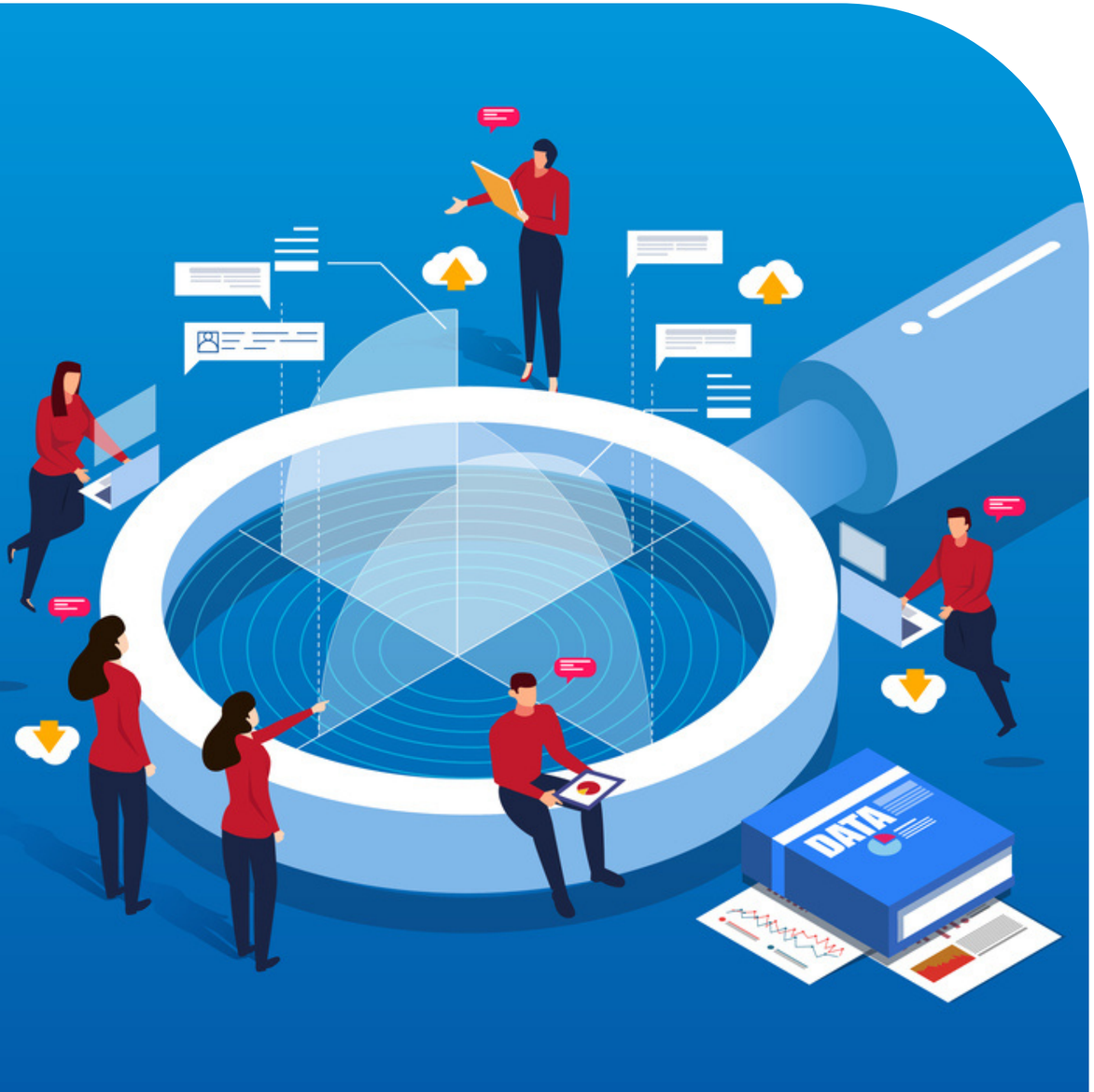
MAPPING SOCIOECONOMIC INDICATORS USING SOCIAL MEDIA ADVERTISING DATA

Presented By
Ahmet Furkan Ün
Enis Mert Kuzu
Taner Giray Sönmez

Scope of the study



- Evaluates the value of publicly accessible advertising data from Facebook.
- Provide strong signals for the Wealth Index (WI) derived from the Demographic and Health Survey (DHS).
- Whether the predictive power of these digital connectivity features.

Research Question

How can anonymous, publicly accessible advertising data be used to map socio-economic development in low and middle-income countries, and how does it provides strong signals for modeling regional variation in the Wealth Index (WI)?

Background Information

According to PEW surveys,

58% of adults in the Philippines,

24% of adults in India

use Facebook.

Materials and Methods

The Demographic and Health Survey(DHS)

- **Survey data**
- **Health and population**

Table 1 Breakdown of the data for each country for clusters with at least one surveyed household

	Philippines	India
Number of DHS clusters	1249	28,524
Clusters missing geo-location	36	131
Geo-located DHS clusters	1213	28,393
Clusters with <100 FB users 18+	8	350
Clusters with ≥100 FB users 18+	**1205**	**28,043**
Clusters with >1000 FB users 18+	1043	25,316
Median number of households surveyed (DHS)	23	21

Materials and Methods

Facebook Data

- **Monthly Active Users**
- **variety of network and device types**
- **18+ Facebook user**

Table 2 List of features derived from the Facebook advertising audience estimate data. All features, with the exception of Facebook penetration, are the fraction of Facebook users in the targeted location who use a given network/device type to access Facebook. All data are for users aged 18+. The Facebook penetration is the number of users divided by the total population of the location; where there were more estimated users than the estimated population the value was capped at 1. Note that according to the Facebook audience estimates, of all users who use a smartphone, the percentage who do not use either of the three specified Mobile OS types (Android, iOS, Windows) are 61% (India) and 51% (Philippines); of all users, the percentage who do not use either of the four specified network types (2G, 3G, 4G, WiFi) to access Facebook are 25% (India) and 37% (Philippines)

Feature type	Feature Description	Correlation with cluster Wealth Index	
		Philippines	India
	Facebook penetration	0.664	0.555
Network access	2G Network	0.115	0.346
	3G Network	–0.378	0.296
	4G Network	0.693	0.003
	WiFi	0.740	0.524
Mobile OS	Android	0.449	0.510
	iOS	0.663	0.567
	Windows phones	0.387	0.357
High-end phones	Apple iPhone X	0.573	0.435
	Apple iPhone X/8/8 Plus	0.628	0.454
	Samsung Galaxy phone S9+	0.540	0.391
	Samsung Galaxy phone S8/S8+/S9/S9+	0.643	0.499
	Samsung Galaxy phone S8/S8+/S9/S9+ or Apple iPhone X/8/8 Plus	0.669	0.524
Other device types	All mobile devices	0.264	–0.061
	Feature phones	0.096	0.163
	Smartphone and tablets	0.217	–0.072
	Tablet	0.492	0.423
	Cherry mobile	–0.275	–
	VIVO mobile devices	0.539	0.024
	Huawei mobile devices	0.534	0.292
	Oppo mobile devices	0.499	0.129
	Oppo/VIVO/Cherry devices	0.184	0.013
	Samsung Android devices	0.123	0.087

Materials and Methods

Population data

- **100 m resolution grid of the entire country for the year 2015.**

Reginional İndicators

- **Binary variables that indicate whether a given DHS cluster falls within a given administrative region in the country**

Models for predicting the Wealth Index

- Linear regression models selected using LASSOs
- Tree-based regression models
- Cross-validation
- RootMean Squared Error (RMSE)

Research Design

- Validity
- Representativeness
- Relevance

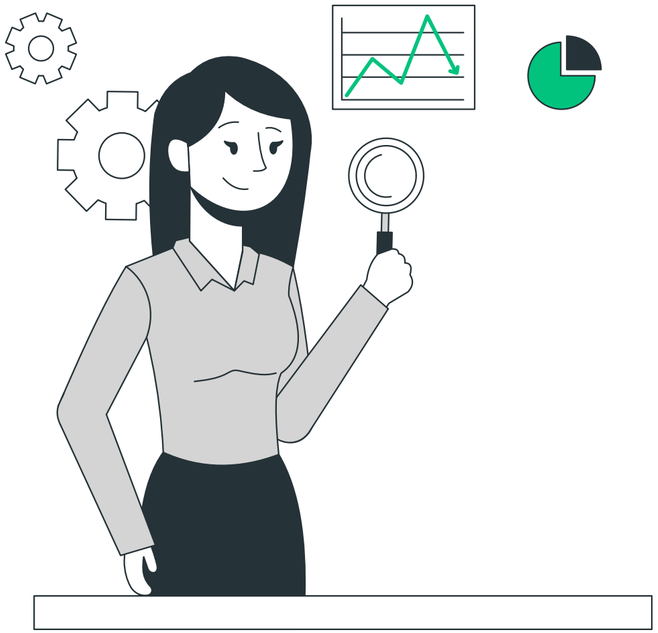
Validity

Does data(e.g Facebook) accurately reflect the true socio-economic situations in these countries?

-Evaluating validity of Facebook data

-Comparing predictions(based on Facebook data) to the actual data

-Demographic and Health Surveys(DHS) actual data(Wealth Index)



Representativeness

Does data(e.g Facebook) represent the whole population in these countries?

-14% of women and 34% of men in India

-Women Facebook users in India -> from the upper socioeconomic strata

-59% of women and 57% of men in the Philippines

Relevance

Does Facebook data relevant and useful in making predictions about the Wealth Index?

-Facebook data is useful -> even in countries with low penetration of Facebook users

-Gives good estimates about Wealth Index

CONCLUSION

- Good for modeling regional variation in the Wealth Index

- Good for predicting Wealth Index

- Can be used to create gender-disaggregated predictions -> but only in gender equal Facebook usage

What are the concerns?

Implications of these concerns?

Can we address this concerns?

Facebook as the main data source for predicting

- Bias towards the ones using Facebook
- May not represent entire population
- Multiple data sources
  - Satellite
    - Electricity consumption at night
    - The use of land

Single Snapshot

- Lack of alignment
- Collect data over a longer period
- Ensure all data is collected at the same time

Wealth Index (WI) as the only measure of socioeconomic development

- May not capture other aspects
  - Access to education
  - Access to healthcare
- Multiple indicators should be used to measure
- Try to include as many aspect as possible

Gender-Disaggregated Prediction

- Especially in contexts with large gender gaps on Facebook
  - India
    - 14% of women
    - 34% of men
    - Biased towards women from the upper socioeconomic strata
- Other sources of information
- Provide equal access

THANK YOU