# A Dataset of Geolocated Villages and Gram Panchayat Election Candidates in Uttar Pradesh

Aryan Srivastava Aarushi Kalra Saket Tiwari

**ABSTRACT**

Village Council or *Gram Panchayat* (henceforth, Panchayat) elections provide highly localized political shocks, making them suitable to answer various research questions in a causal framework. We collect data for Panchayat elections held in the Indian state of Uttar Pradesh for the years 2021 and 2015. While election candidates' data are openly available on Uttar Pradesh State Election Commision website, they are not easily accessible. Moreover, the villages are not geolocated. This motivated us to scrape the state website to obtain the data, and geolocate it to allow us to match them with our social media data consisting of the user's activity and their geo-location. These data are a valuable resource for researchers interested in questions related to political economy in the Developing World. Link to the Github repository with the code and data can be found here[1].

[1] https://github.com/aryan096/Uttar-Pradesh-2015-2021-Panchayat-Candidates

## I. MOTIVATION

*Gram Panchayat* (henceforth, Panchayat) elections are local council elections in rural India. These elections consist of 500 to 10,000 adults voting to elect the representatives in around 60,000 villages in the state of Uttar Pradesh alone (Indian Census, 2011 [7]). These provide highly localized political shocks, making them suitable to answer various research questions in a causal framework. Suppose researchers are interested in analyzing user engagement with hateful content on social media when a muslim political candidate won panchayat elections over an incumbent non-muslim leader, and vice versa. We may want to interrogate if the election of Muslim leaders leads to backlash against minority communities (Mitra and Ray, 2014 [8]) in the form of hate speech on social media.

While election candidates' data is openly available on Uttar Pradesh State Election Commision website, it is not easily accessible. Moreover, the villages are not geolocated. This motivated us to scrape the state website to obtain the data, and geolocate it to allow us to match them with our social media data consisting of the user's activity and their geo-location.

These data are a valuable resource for researchers interested in questions related to political economy in the Developing World. Data on Panchayat elections, as well as identity of local political leaders (known as *Gram Pradhans*) have enabled research on effects of gender representation in local politics (Chattopadhyay and Duflo, 2004 [5]; Beaman et al, 2009 [2]), consequences of local democratization on the distribution of public goods (Foster and Rosenzweig, 2001 [6]; Besley et al, 2004 [3]; Sharan, 2020 [9]), and has enabled deeper understanding of local institutional structure (Bardhan et al, 2010 [1]; Bhattacharya et al, 2016 [4]).

## II. COMPOSITION & COLLECTION

Candidate data was scraped from the Uttar Pradesh State Election Commission[2] website using Selenium in Python. 2021 election data was then joined with 2015 election data using both the *Zila* and *Gram Panchayat* name. Panchayats were then geolocated using the Google Maps API. There is data for 52,880 villages across Uttar Pradesh. Only villages that held panchayat elections in 2015 and 2021 are included in the dataset. Using Google Maps API, we geolocated the villages with ~65% accuracy. We achieved a maximum of ~25% accuracy with other approaches like matching village names to other datasets[3] containing geolocations. Table 1 contains a description of the variables in the dataset.

We also classified the religion of each candidate as muslim or non-muslim using a neural network model with 95% accuracy. This was created using a large dataset of 15,000 Hindi names annotated as Muslim or Non-Muslim. We are neither releasing the classifier nor the classifications to the public. Subpopulations can be recognized in the dataset by age, gender, caste, education, and location.

---

[2] 2021 Elections - http://sec.up.nic.in/site/DownloadCandidateFaDebt.aspx
2015 Elections - http://sec.up.nic.in/ElecLive/WinnerList.aspx

[3] OpenStreetMap - http://download.openstreetmap.fr/extracts/asia/india/
SEDAC Census 2001 - https://sedac.ciesin.columbia.edu/data/set/india-india-village-level-geospatial-socio-econ-1991-2001
ML Infomap - https://www.mlinfomap.com/

| | |
|---|---|
| id | Index number, starting from 0 |
| matching | zila + gram_panchayat. unique name used for matching |
| zila | name of the zila (district) |
| vikas_khand | name of the vikas khand (block) |
| gram_panchayat | name of the gram panchayat (village) |
| winner_name_2021 | name of winner - gram panchayat pradhan elections 2021 |
| winner_father_husband_name_2021 | name of winner's father or husband - gram panchayat pradhan elections 2021 |
| winner_caste_2021 | caste of winner - gram panchayat pradhan elections 2021 |
| winner_education_2021 | highest level of education of winner - gram panchayat pradhan elections 2021 |
| winner_gender_2021 | gender of winner - gram panchayat pradhan elections 2021 |
| winner_name_2015 | name of winner - gram panchayat pradhan elections 2015 |
| winner_father_husband_name_2015 | name of winner's father or husband - gram panchayat pradhan elections 2015 |
| caste_2015 | caste of winner - gram panchayat pradhan elections 2015 |
| winner_education_2015 | highest level of education of winner - gram panchayat pradhan elections 2015 |
| winner_gender_2015 | gender of winner - gram panchayat pradhan elections 2015 |
| eng_matching | English transliteration of the matching column |
| latitude | Latitude of Google Maps location |
| longitude | Longitude of Google Maps Location |
| runner_up_name_2021 | name of runner up - gram panchayat pradhan elections 2021 |
| runner_up_father_husband_name_2021 | name of runner up's father or husband - gram panchayat pradhan elections 2021 |
| runner_up_vote_percentage_2021 | Vote percentage of runner up - gram panchayat pradhan elections 2021 |
| winner_vote_percentage_2021 | Vote percentage of winner - gram panchayat pradhan elections 2021 |

**Table 1. Description of variables in the dataset**

## III. USE, DISTRIBUTION, AND MAINTENANCE

This dataset is openly available for people to view, download, and use upon appropriate citation. The .csv file can be downloaded using the Google Sheets link in the Github repository. We have neither released religion classifications nor the religion classifier. Please reach out to the authors if you are interested in using religion data. While this dataset only contains *Gram Panchayat Pradhan* election data, data about other state elections (eg: *zila panchayat* and *kshetra panchayat* elections) can be scraped by simple modifications to our code.

## IV. REFERENCES


[1] Bardhan, Pranab K., Dilip Mookherjee, and Monica Parra Torrado. "Impact of political reservations in West Bengal local governments on anti-poverty targeting." *Journal of Globalization and development* 1, no. 1 (2010).

[2] Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova. "Powerful women: does exposure reduce bias?." *The Quarterly journal of economics* 124, no. 4 (2009): 1497-1540.

[3] Besley, Timothy, Rohini Pande, Lupin Rahman, and Vijayendra Rao. "The politics of public good provision: Evidence from Indian local governments." *Journal of the European Economic Association* 2, no. 2-3 (2004): 416-426.

[4] Bhattacharya, Anindya, Anirban Kar, and Alita Nandi. "Local institutional structure and clientelistic access to employment: the case of MGNREGS in three states of India." (2016).

[5] Chattopadhyay, Raghabendra, and Esther Duflo. "Women as policy makers: Evidence from a randomized policy experiment in India." *Econometrica* 72, no. 5 (2004): 1409-1443.

[6] Foster, Andrew D., and Mark R. Rosenzweig. "Democratization, decentralization and the distribution of local public goods in a poor rural economy." (2001).

[7] "Indian Census". Censusindia.gov.in. Retrieved 9 April 2012.

[8] Mitra, Anirban, and Debraj Ray. "Implications of an economic theory of conflict: Hindu-Muslim violence in India." *Journal of Political Economy* 122, no. 4 (2014): 719-765.

[9] Sharan, M. R. "It's Complicated: The Distributional Consequences of Political Reservation."