DIAGNOSING WEB DATA OF ICTS TO PROVIDE FOCUSED ASSISTANCE IN AGRICULTURAL ADOPTIONS

Ashwin Singh[1], **Mallika Subramanian**[1], **Anmol Agarwal**[1], **Pratyush Priyadarshi**[1], **Shrey Gupta**[1]

Kiran Garimella[2], **Sanjeev Kumar**[3], **Ritesh Kumar**[3], **Lokesh Garg**[3], **Erica Arya**[3]

Ponnurangam Kumaraguru[1]

[1] International Institute of Information Technology, Hyderabad, India
[2] MIT Institute for Data, Systems, and Society
[3] Digital Green, India
ashwin19.iiith@gmail.com
{mallika.subramanian, anmol.agarwal, pratyush.priyadarshi, shrey.gupta}@students.iiit.ac.in
garimell@mit.edu
{sanjeev, ritesh, lokesh, erica}@digitalgreen.org
pk.guru@iiit.ac.in

ABSTRACT

The past decade has witnessed a rapid increase in technology ownership across the rural areas of India, signifying the potential for ICT initiatives to empower rural households. In our work, we focus on the web infrastructure of one such ICT - Digital Green, a global development organization that started in 2008. Following a participatory approach for content production, Digital Green disseminates instructional agricultural videos to smallholder farmers via human mediators to improve the adoption of farming practices. Their web-based data tracker, CoCo, captures data related to these processes, storing the attendance and adoption logs of over 2.3 million farmers across three continents and twelve countries. Using this data, we model the salient components of the Digital Green ecosystem involving the past attendance and adoption behaviours of farmers, content from the videos screened to them and the demographic features of farmers across five states in India. We use statistical tests to identify different factors which distinguish farmers with higher adoption rates to understand why they adopt more than others. Our research finds that farmers with higher adoption rates adopt videos of shorter duration and belong to smaller villages. The co-attendance and co-adoption networks of farmers indicate that farmers greatly benefit from past adopters of a video from their village and group when it comes to adopting practices from the same video. Following our analysis, we model the adoption of practices from a video as a prediction problem to identify and assist farmers who might face challenges in adoption in each of the five states. We experiment with different model architectures and achieve macro-f1 scores ranging from 79% to 89% using a Random Forest classifier. Finally, we measure the importance of different components of the Digital Green ecosystem using SHAP values and provide implications for improving the adoption rates of nearly a million farmers across five states in India.

Keywords Web Diagnosis · ICT4D · Agriculture · Social Networks

1 Introduction

Agriculture is the primary source of livelihood for 58% of India's population [1]. The agriculture sector accounted for 20.2% of the country's GDP in 2020-21, ranking third, only below the services and industry sectors. [1] Over the last 20 years, India has prevailed as the seventh-largest agricultural exporter worldwide, with the sector inviting foreign direct investment of over $10.24 billion [1]. As of September 2021, there are 17.16 million farmers who are registered with the Electronic National Agricultural Market (e-NAM) [2], demonstrating the proliferation of technology in rural areas of India and signifying potential for ICT-based initiatives to empower rural households. Digital Green is one such ICT initiative that aims to empower smallholder farmers by leveraging technology and grassroots-level partnerships. [3] They disseminate instructional content through the medium of videos to rural populations in India and other countries, with the goal to improve the adoption of practices related to agriculture, health and livestock alongside raising awareness about topics such as social issues and financial management.

In our work, we focus on the web infrastructure used by Digital Green called CoCo (Connect Online Connect Offline), which is used to capture data related to their key processes i.e., video production, dissemination and adoption of practices. [4] We also limit our scope to only the agricultural practices as the adoption mechanisms for other types of practices can be inherently different. In the next two subsections, we cover the relevant background about Digital Green, ICTs for rural development and adoption of agricultural practices to motivate our research questions.

1.1 Digital Green and ICT 4 Rural Development

Digital Green is built upon two salient features involving its human infrastructure [2]. First, they follow a participatory process for video-content production to tailor it for the local communities. Second, they make use of human-mediated instruction to engage the rural people in the dissemination and training process. We describe this process in more detail in Section 2; This idea of involving human actors in mediating instructions was later recognised as '*infomediaries*' in the ICT4D 2.0 Manifesto [3]. While early rural ICTs in India such as IShakti facilitated the access of personalised advice with domain experts [4] to underserved and isolated communities, they lacked information specificity sought out by farmers. To overcome this limitation, ICTs have seen a rise in participatory video production, which allows for the representation of local communities and tapping into infomediaries to disseminate relevant information [2]. Similar strategies have been observed across health workers in rural India [5], where ASHAs (Accredited Social Health Activists) are actively involved in engaging high-status infomediaries in the process of video production.

Digital Green's initial study in 2007 involved a four-month trial across 16 villages in India, which saw an increase in the adoption of specific agriculture practices by a factor of six-seven times over traditional modes of television programs and radio broadcasts [2]. As of 2021, Digital Green has scaled across three continents and twelve countries, reaching 2.3M rural households globally. [5] It has been particularly lauded by experienced HCI4D researchers [6] in terms of scaling across continents and standing the test of time, even with the evolution in technologies. It has been funded by USAID, World Bank, The ICCO Cooperation [6] and the Bill & Melinda Gates Foundation in recent years. [7] Digital Green has also served as a model example of an ICT where the human infrastructure is central to supporting knowledge transfer via a digital medium [7]. Over the years, it has introduced several extensions, a notable one being 'Videokheti' [8, 9], which improved the accessibility of videos to farmers by allowing them to rewatch screened videos. To further increase this accessibility, Digital Green has also established its presence on YouTube with over 287K subscribers [8] and curated a digital library of videos on its website. [9]

1.2 Big Data and Adoption in Agriculture

Understanding the dynamics of components that influence adoptions of agricultural practices can greatly benefit farming communities [10, 11, 12]. Broadly, these components can be associated with environmental factors, institutional structures, government influence and the information flow dynamics within a community [13]. Across each of these, there is potential for big data production in the agriculture industry. However, as highlighted by Kamilaris et al. [14], the current sources of big data in agriculture are limited to remote and proximal sensing tools, historical records of food and climate data, static databases of geospatial data, surveys conducted by the government and web-based accounts of farmers' decision-making. First, while some of these sources enable studying macro-level socio-economic and policy indicators at scale [15, 16, 17], they lack consideration for the social dynamics and ground-level interactions that unfold between farming communities. Second,

[1] https://www.pib.gov.in/PressReleasePage.aspx?PRID=1741942
[2] https://enam.gov.in/web/dashboard/stakeholder-data
[3] https://www.digitalgreen.org/about-us/
[4] https://www.digitalgreen.org/coco/
[5] https://www.digitalgreen.org/global-impact/
[6] https://www.digitalgreen.org/global-initiatives/
[7] https://www.gatesfoundation.org/about/committed-grants/2020/10/inv004995
[8] https://www.youtube.com/user/digitalgreenorg/
[9] https://solutions.digitalgreen.org/videos/library

proximal sensing sources enable IoT and cloud-based innovations that support farmers in agriculture [14, 18] but are dependent on technology ownership and thereby susceptible to deepening the digital divide [19, 20]. Third, there is a lack of big data when it comes to capturing the social dynamics of farmers. One can model them as social networks to study knowledge sharing. While previous works [21, 22, 23] look at various ways of information diffusion in such networks, the scale of their evaluation is limited to small and localised farmer populations. Lastly, when it comes to predicting and analysing trends in the adoption of farming practices, past works have been limited to specific practices [24, 25], highlighting the scope for studying how factors impact a multitude of farming practices.

1.3 Research Questions and Contributions

CoCo enables us to overcome the limitation of social data at scale by capturing the data of over 1.9M farmers in India involving their attendance and adoption logs. It also consists of over 2.7K videos containing instructional content, allowing us to study diverse agricultural practices. We perform a holistic diagnosis for the ICT using CoCo, looking at how various social, temporal and content features influence adoption. In particular, we ask the following research questions:

- What are the differential factors for farmers which impact the adoption of agricultural practices from a video?
- How important are these factors, and how does their importance vary for farmers across different states?
- How can we identify farmers who face challenges in adopting farming practices to provide assistance to them?

First, our analysis elaborates on the different factors where farmers with relatively lower adoption rates require alleviation and assistance in adopting farming practices across different states. Second, it helps Digital Green in identifying them in order to provide focused assistance and enable further qualitative investigation into their experiences. Third, it also assists them in making decisions concerning the screening of videos across various parameters such as their duration and content. Lastly, it detects gender-based inequalities in Digital Green so that they can be mitigated.

2 The Digital Green Ecosystem and Dataset Description

The Digital Green (DG) ecosystem consists of various actors and components, which are described in Table 1. It starts with the participatory production of video content, where content producers (scientists, NGO experts, field staff and progressive farmers) involve the local farmers in creating instructional videos tailored for the community [2]. Mediators with varying levels of expertise (frontline workers and extension officers) conduct screenings to disseminate these videos to groups of farmers from the local community. Mediators are also supported by the partners who are employees from the government or NGOs such as Bharatiya Agro Industries Foundation [10] and Samaj Pragati Sahayog. [11] These partners enable feedback and audit mechanisms for clusters of villages. Approximately two weeks after a screening is conducted, the staff associated with DG go on-site to survey farmers along with the mediator and associated partner to verify the adoption of practices disseminated in the videos. A video consists of three to five key recall points corresponding to each practice, which are either verified physically or through knowledge transfer by the surveyor. These surveyors report back to centres and the data entry operators input this adoption data into CoCo.

Table 1: Terminology for Actors and Components

Actor/Component	Description
Farmer	A person pursuing farming as a member of the Digital Green Ecosystem
Group	Self Help Groups of Farmers formed by the Government
Partner	An NGO or government organisation associated with the activities of DG.
Mediator	A Frontline Worker or Officer who disseminates the videos to Farmers
Video	A Video containing practices relevant to the Farmers
Screening	Screening of relevant video(s) via a projector to target Groups of Farmers
Adoption	Verified instances of Farmers implementing or learning practices from the video

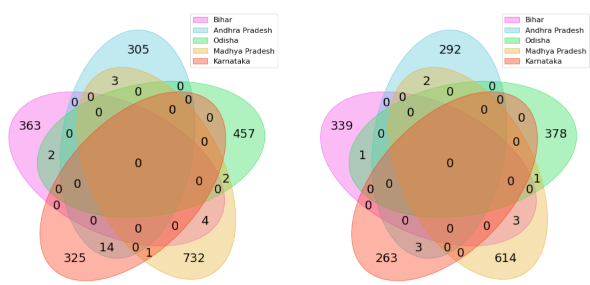
CoCo contains the data for agricultural screenings and adoptions of videos across a period of ten years between 2010 and 2020. For our work, considering the page limitation, we focus on only the top five of twelve states in India where DG is most active in terms of the number of screenings conducted – Bihar, Odisha, Andhra Pradesh, Karnataka and Madhya Pradesh. These states are divided into districts which are further divided into blocks that are constituted by villages at the lowest level. A comprehensive view of the descriptive statistics such as the number of videos, screenings, adoptions, farmers and geographic distribution for these five states is presented in Table 2.

[10] https://baif.org.in/
[11] http://www.samajpragatisahayog.org/

Table 2: Dataset Description and Statistics, showing the scale of the CoCo Web Infrastructure. (*) Numbers only represent the unique videos screened and adopted. A video can be screened and adopted multiple times and across different states as shown in the Venn Diagram (see Figure 1).

State	Districts	Blocks	Villages	Groups	Farmers	Mediators	Screenings	Videos Screened	Videos Adopted
Bihar	38	243	4,908	46,621	534,507	5,021	232,994	369	343
Andhra Pradesh	17	288	2,951	18,274	221,052	2,409	90,163	322	297
Odisha	7	32	1,262	6,773	107,665	553	78,461	461	380
Madhya Pradesh	17	72	1,295	4,480	67,478	801	57,739	742	620
Karnataka	15	35	780	4,074	50,923	624	32,719	340	266
Total	94	670	11,196	80,222	981,625	9,408	259,082	2,208*	1,896*

One of the salient features of DG is its participatory approach to video production i.e., their content is highly tailored to the local communities involved in its production. Figure 1 shows a Venn diagram of the unique videos screened and adopted across the five states. We observe that most of these videos are specific to the states and the adoption of videos is particularly low at their intersections, highlighting their local community-based approach.



(a) Videos Screened (b) Videos Adopted

Figure 1: Venn Diagram of (a) videos screened and (b) videos adopted across the five states. Of 461 videos in Odisha, only two were screened in Bihar and Madhya Pradesh each, and only one was adopted in each state, demonstrating high content specificity and DG's community-based approach.

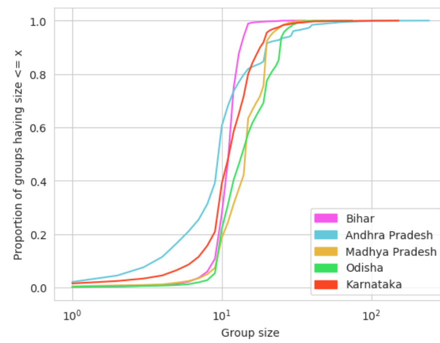


Figure 2: CDF Plot for Group Sizes (μ=12.24, σ=6.05). Eighty one percent of the Groups comprise 10-30 farmers to ensure a healthy mediator to farmer ratio.

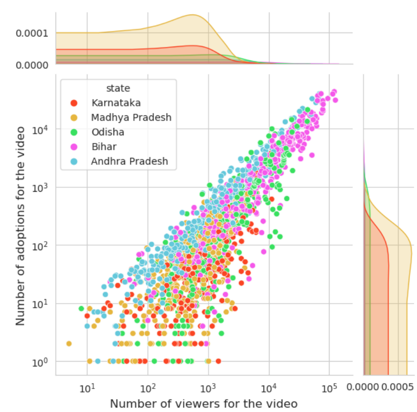


Figure 3: Joint plot capturing distribution and log-linear trend of adoptions with number of views of a video. Fewer videos are widely adopted as visible from the distribution and funnel shape of the plot.

To quantify it further, we look at this specificity at the village level and find that out of all pairs of villages in each state, the percentage of village pairs adopting at least one common video is very low – Bihar (9.6%), Andhra Pradesh (26.4%), Odisha (11.3%), Madhya Pradesh (12.1%) and Karnataka (8.3%). This demonstrates the high specificity of DG's content across the five states, even at the lowest geographical level. To disseminate this specific content, farmers are divided into self help groups as beneficiaries of government schemes. These groups attend screenings of the same videos together. There are a total of 80,222 (Column 5, Table 2) such groups across the five states with varying sizes (μ = 12.24, median = 12, σ = 6.05). The CDF plot (Figure 2) of the group sizes for each of the five states depicts that a large percentage of the farmer groups (81.74%) comprise 10-30 farmers.

We examine the videos closely in Figure 3 and find that the number of adoptions for a video follows a log-linear trend with its number of views because only a fraction of the viewers adopt the video. The videos in the plot funnel towards the end, denoting that there are only a few videos that are widely adopted. We investigate the video-screening and adoption behaviours further by looking at temporal patterns. In Figure 4, we plot the time-series trends for both behaviours. We notice that the spikes in adoptions of farmers almost coincide with the spikes in the screenings of videos across a period of ten years.

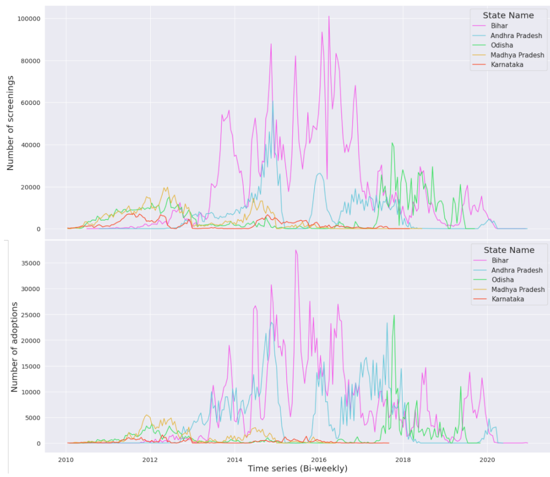


Figure 4: Timeseries trends for screenings and adoptions. The spikes in adoptions of farmers almost coincide with the spikes in the screenings of videos, depicting how the adoption behaviour of farmers loosely mimics the trends in screening by DG.

3 Modelling Components of the Digital Green Ecosystem

In this section, we motivate and model relevant features corresponding to the three components present in CoCo – (i) attendance and adoption behaviours of farmers, (ii) content details of the videos screened to farmers, and (iii) demographic features of farmers and mediators.

3.1 Attendance and Adoption Behaviours

We model the attendance and adoption behaviours of farmers using two temporal networks – (i) $G_1 = (F, E)$ where F denotes the set of nodes (farmers) and E denotes the set of edges where an edge $(f, g, w, d) \in E$ represents two farmers $f, g \in F$ who have co-attended w screenings of agricultural videos prior to date d, and (ii) $G_2 = (F, E)$ where an edge $(f, g, w, d) \in E$ represents two farmers $f, g \in V$ who have co-adopted agricultural practices from w videos before date d. Due to the high specificity of content in CoCo as seen in Section 2, we restrict the set of nodes F to farmers belonging to the same village i.e., we construct G_1, G_2 for all villages across the five states to capture the attendance and adoption dynamics of people residing there. For each farmer, we compute three centrality measures for both G_1 and G_2 – **Closeness** (**CC**), **Betweenness** (BC) and **Eigenvector** (EC) to consider how different farmers fare based on their position in their networks. We compute all three centrality measures temporally i.e., for a farmer $f \in F$ watching the screening of a video v on date d, we only consider the edges in temporal networks G_1 and G_2 upto date d. Next, for a video v being screened to a farmer f on date d, we consider farmers $g \in N(f, d)$ in G_2 who have adopted the video v before date d to measure how past adopters of the same video from the neighbourhood of f in G_2 can influence adoption of a video by farmer f. We measure this **Past Co-Adopter Influence (PAI)** at two levels – the village and group of farmer f and formulate it as follows:

$$PAI_L(f, v, d) = |\ N(f, d) \cap A_L(v, d)\ | \qquad (1)$$

where L represents level and $A_L(v, d)$ denotes the adopters of video v before date d at level L. Lastly, for a farmer f watching the screening of a video v on date d, the attention given to them by the mediator during video dissemination can vary depending on the number of co-attendees $|A_{v,d}|$. We formulate this **Mediator Attention (MA)** as follows:

$$MA(f, v, d) = \frac{1}{|A_{v,d}|} \qquad (2)$$

3.2 Content Features

We extract the following content-based features from each video to capture relevant information for our model:

i) Duration: To account for the farmers' attention span and understand how different duration lengths of videos help people assimilate information, we include the duration of a video (in minutes) as a feature.

ii) Language: To account for the linguistic diversity of viewers in our data, we represent the 19 Language IDs in our database as One-Hot Vectors.

iii) Content Specificity (CS): As seen in Section 2, the videos in CoCo have a high content specificity due to DG's participatory approach. We model this as:

$$CS_L(video) = \frac{1}{\sum_{v \in VS(L)} |L|} \qquad (3)$$

where v denotes video, $L \in \{village, block, district\}$ denotes level, $VS(L)$ represents the set of videos screened at level L and $|L|$ is the population of the level.

iv) Title Adoption TF-IDF (TA): Some video topics are more prevalent among farmers than others. We measure this temporally for each video, date pair (v, d). We compute the cumulative sum of the adoptions per word in the title of video v till date d and normalize it by the number of screenings of v till date d:

$$TA(v, d) = \frac{\sum_{word \in Title(v)} A(word, d)}{S(v, d)} \qquad (4)$$

where $A(word, d)$ = number of adoptions of the word across all video, date pairs in a state till date d, and $S(v, d)$ = number of screenings of v before d.

v) Time of Screening: Screening of videos can be conducted during different times of the day. To account for the preferences of farmers, we divide these into bins of four hours each throughout the day starting 4 am – early morning, morning, noon, evening, night, late night before encoding them as one-hot vectors.

3.3 Demographic Features of Actors

For our model, we first consider a farmer's **group size** and **village size** to consider the extent to which the size of their community impacts their adoptions. Then for each farmer f viewing a video v on date d, we measure their **active age** as the number of days between their first screening and d to consider the duration of their association with DG temporally. In section 4.2, we make use of the gender of farmers and mediators to diagnose the Digital Green ecosystem for potential inequalities.

4 Understanding Differential Factors in Adoption

In this section we try to understand why some farmers adopt more videos than others based on the various factors that govern the differences between them. To study, this we define the **adoption rate (AR)** for each farmer as follows:

$$AR(farmer) = \frac{|\ videos\ adopted\ by\ farmer\ |}{|\ videos\ viewed\ by\ farmer\ |} \quad (5)$$

We plot the CDF of adoption rate for farmers across the five states (see Figure 5). We observe that a significant percentage of the farmers have not adopted any videos - Karnataka (74%), Odisha (49%), Madhya Pradesh (49%), Bihar (48%), and Andhra Pradesh (34%).

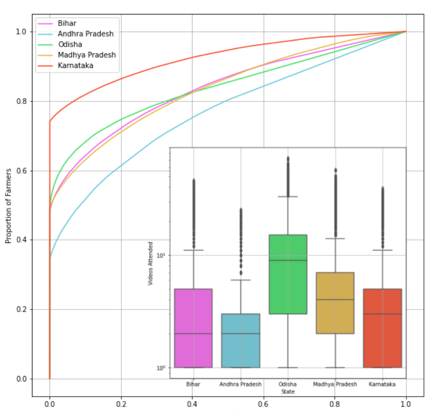


Figure 5: CDF Plot for Adoption Rates of Farmers across the five states showing that a large percentage of Farmers have no adoptions. Inset Plot represents box plot of videos attended by farmers with AR = 0.

On average, these farmers have attended the screening of fewer videos than other farmers who have adopted at least one video, with their statewise means ($\mu_{AR=0} < \mu_{AR>0}$) – Bihar (3.81 < 9.98), Andhra Pradesh (2.67 < 5.36), Odisha (10.22 < 12.17), Madhya Pradesh (5.23 < 14.71) and Karnataka (4.1 < 9.73). We speculate their reasons to be attributed to their socio-economic status, lack of resources or them adopting sub-practices that do not fulfil the surveyor's criteria for adoption. However, we plan to conduct an ethnographic investigation in the future to better understand these reasons since our data doesn't account for their lived experiences. For the scope of our analysis, we only consider the farmers with at least one adoption (AR > 0).

4.1 Why Some Farmers Adopt more than Others

We divide the farmers across each of the five states into quartiles based on their adoption rates to understand how the lowest 25% (q_1) and top 25% (q_4) farmers vary in terms of factors specific to all farmer, video (f, v) pairs. To do so, we consider eight factors – mediator attention (MA), content specificity (CS), past co-adopter influence (PAI) at the village and group level, video duration, group size (GS) and village size (VS) and active age. For the first five factors, we consider the mean value across all videos attended by each farmer. We make use of one-tailed Welch's *t*-test to evaluate our hypotheses across all the factors (Table 3). For MA_{μ}, $CS_{V\mu}$, $PAI_{G\mu}$, $PAI_{V\mu}$ and $Active\ Age$, we test the hypothesis $H_1 : q_1 < q_4$. We evaluate if higher mediator attention, content specificity at the village level, past co-adopter influence at group-village levels and longer active association with DG result in higher adoption rates. For the other three, we test the hypothesis $H_1 : q_4 < q_1$ to evaluate whether longer duration of videos ($duration_{\mu}$), and larger sizes of villages (VS) and groups (GS) lead to lower adoption rates. Given that we test eight different hypotheses using the same samples, we apply the Bonferroni correction (number of measures *m=8*) to the p values while considering statistical significance. We only report the results for $\alpha = 0.001/m$. Mean values for q_1 and q_4 across all the eight factors are reported in the Appendix Section A.

First, we infer that farmers with higher adoption rates (q_4) watch videos of shorter duration and belong to smaller villages across all the five states as compared to the farmers in q_1. Second, for all states except Andhra Pradesh, farmers in q_4 highly benefit from farmers belonging to their neighbourhood in G_2 who are adopters of a video that is now being attended by the farmer. Third, farmers in q_4 from Andhra Pradesh and Karnataka watch videos that are more specific to their villages as compared to q_1. This tells us that since farmers in q_1 from Karnataka watch videos that are less specific to their villages, they might be less likely to benefit from other farmers in their group/village adopting the same videos in the past. Fourth, farmers in q_4 from Bihar, Andhra Pradesh and Madhya Pradesh watch videos with lower attendance, benefitting from a higher mediator attention (MA_{μ}) as compared to farmers in q_1. Fifth, farmers in q_4 from Bihar and Madhya Pradesh belong to smaller groups as compared to q_1. Lastly, the mean active age of farmers in q_4 was lesser than farmers in q_1 (see Appendix Section A) in contrast to our hypothesis. We delve into this in more detail in Section 5.2.

Table 3: One-Tailed Welch t-Tests for eight factors between farmers in q_1 and q_4 of adoption rates. t-stat is reported only in cells where $\alpha = 0.001$ after adjusting p values as per the Bonferroni correction (number of measures $m = 8$).

	$q_1 < q_4$					$q_4 < q_1$		
State	MA_μ	$CS_{V\mu}$	$PAI_{G\mu}$	$PAI_{V\mu}$	Active Age	duration $_\mu$	GS	VS
Bihar	-43.22	–	-9.91	-51.47	–	-12.89	-54.89	-23.81
Andhra Pradesh	-21.96	-39.66	–	–	–	-6.13	–	-21.22
Odisha	–	–	-24.82	-23.81	–	-87.03	–	-33.93
Madhya Pradesh	-26.12	–	-15.75	-16.51	–	-10.03	-31.84	-12.30
Karnataka	–	-16.39	-6.64	-3.78	–	-20.12	–	-35.23

4.2 Gender-Based Inequalities

We make use of the gender of farmers and mediators in our data to diagnose the Digital Green ecosystem for inequalities. The distribution of genders for both farmers and mediators across the five states is shown in Figure 6. We observe that majority of the farmers in all the states except Madhya Pradesh are women whereas the proportion of men is higher for mediators across all states.

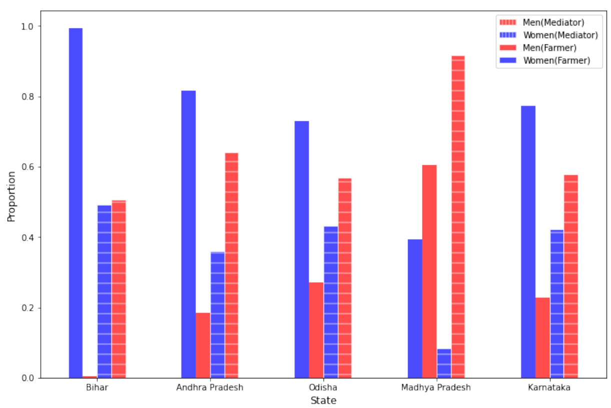


Figure 6: Bar Plot showing the Proportion of Farmer and Mediator Genders across the five states. Majority farmers in all states except Madhya Pradesh are Women whereas the proportion of Men is higher for mediators across all states.

First, we define the adoption rate (AR) for a mediator as follows:

$$AR(mediator) = \frac{\sum_{v \in V(mediator)} \frac{|\ adoptions\ of\ v\ |}{|\ attendees\ of\ v\ |}}{|V(mediator)|} \tag{6}$$

where $V(mediator)$ denotes the set of videos disseminated by the mediator to the farmers. We only consider attendees and adoptions for the screening conducted by the mediator.

Digital Green aims to empower smallholder and marginalised farmers in villages across India, most of who are women. [12] Therefore, we test the hypothesis $H_1 : AR_\mu(Women) < AR_\mu(Men)$ for both farmers and mediators using a one-sided Welch's t-test across the five states to facilitate our diagnosis. We report results for two significance levels $\alpha = 0.05$ and $\alpha = 0.001$ in Table 4.

Table 4: One-Tailed Welch's t-test for Farmer and Mediator Gender. We report t-stat for $\alpha = 0.05$(*), 0.001(**)

(a) One-Tailed Welch's t-test for Farmer Gender

State \ **Farmer**	$AR_\mu(M)$	$AR_\mu(W)$	***t*-stat**
Bihar	0.2322	0.4589	-
Andhra Pradesh	0.6871	0.5737	-60.31**
Odisha	0.7150	0.5207	-57.89**
Madhya Pradesh	0.2565	0.2810	-
Karnataka	0.3712	0.2420	-20.02**

(b) One-Tailed Welch's t-test for Mediator Gender

State \ **Mediator**	$AR_\mu(M)$	$AR_\mu(W)$	***t*-stat**
Bihar	0.2346	0.2210	-1.83*
Andhra Pradesh	0.4014	0.3928	-
Odisha	0.4115	0.2043	-
Madhya Pradesh	0.1566	0.1139	-2.09*
Karnataka	0.0696	0.0862	-

For farmers, the Welch's t-test (Table 4a) informs that the disparities in the adoption rates of men and women are very highly significant in Andhra Pradesh, Odisha and Karnataka (greater t-stat denotes more disparity). For mediators, the t-test (Table 4b) highlights that men are more effective mediators in Bihar and Madhya Pradesh.

[12] https://www.digitalgreen.org/india/

5 Predicting Adoption of Agricultural Practices

In the previous section, we learned how farmers with lower AR (q_1) differ from those with higher AR (q_4). One of our main objectives is to assist DG in improving the adoption rates for such farmers (q_1). Therefore, to identify farmers who are less likely to adopt practices from a video screened on a particular date, we model this problem as a prediction task. The complete pipeline of adoptions, starting from preparing the content for a video, disseminating the practices via screenings and finally the farmer adopting the practice, involves several key components which we defined in Section 3. We leverage them for our model and explain how various features impact its output differently across the five states. A prior estimate about the response of a video being screened will enable DG to support the farmers who are not likely to adopt the video and conduct on-field investigations to better understand their reasons. Therefore, we try to predict whether a farmer f will adopt a video v being screened to them on date d. The next section describes our model setup in detail.

Table 5: Class Distribution in data across the five states

State \ Class	No Adoption	Adoption
Bihar	2,180,336	1,188,162
Andhra Pradesh	514,456	549,094
Odisha	490,386	239,687
Madhya Pradesh	439,639	149,599
Karnataka	138,891	35,770

5.1 Model Setup:

In Section 2, we learned that the videos of Digital Green is highly specific to the states they screen them in. In Section 4, while understanding differential factors in adoption, we again observed most of the trends being distinct to the states. Hence, we acknowledge this diversity and divide our data across the five states based on the location where the screening is conducted. In Figure 4, we notice that the number of adoptions for each state has increased more in recent years as compared to the initial years of DG, given that they have witnessed tremendous growth since 2015. All our features have been computed temporally for each (f, v, d) triplet in this timeline. Hence, we utilize the conventional methods to split the train-test data for our model. The distribution of both classes (No Adoption: 0, Adoption: 1) for all the (f, v, d) pairs in the five states can be observed in Table 5. To overcome the class imbalance, we perform down sampling for the majority class in the preparation of our train set. We train five models, one for each of the five states. We experiment three different classification techniques to predict adoptions – (i) Logistic Regression, (ii) XGBoost *(boosting stages = 25, lr = 0.1)* and (iii) Random Forest (*trees = 25, depth = max*).

5.2 Model Results and Explainability

For evaluating our model, we use two metrics – (i) True Negative (TN) Rate which is important in identifying farmers facing challenges in adopting practices, and (ii) the macro-f1 score to account for the class-imbalance in the dataset while explaining feature importance. The performance of all three models across the five states is summarized in Table 6. We find that the Random Forest Classifier outperforms the other two models on both the metrics.

We use Shapley Additive Explanations, or SHAP values [26] to measure the feature importance. These explanations capture the contribution of each feature in the model based on local explanations [27]. Therefore, we produce SHAP plots for class 1 (Adoption) (see Figure 7) for every state to measure and explain the impact of different features in predicting the adoption of farming practices.

First, we note that PAI at the group-village levels positively impacts the adoption of farming practices as per our model. This hints at the role played by past adopters of the same video who belong to the farmer's community in transferring relevant knowledge. Second, influential farmers (high EC) and farmers connecting multiple communities (high BC) in the co-adoption network are more likely to adopt practices from a video than others. Third, across all states except Karnataka, farmers best placed to be influenced by the network (high CC) also have a higher likelihood of adopting practices as per our model. Fourth, we find that longer videos have a negative impact on adoption across all states except Bihar, i.e., shorter videos are preferred by the farmers. We verify this further by fitting a regression line onto the SHAP dependency plots across all states (see Appendix Section B) to find that only Bihar has a positive slope with duration. Fifth, for content specificity (CS), we observe that it mostly has a negative or mixed impact on adoptions. [13] The SHAP indicates its positive impact only for the district and block levels in Karnataka. We verify its negative impact across rest of the states by fitting a regression line onto the SHAP dependency plots of content specificity on the three levels (see Appendix Section B). Sixth, the title adoption TF-IDF positively impacts adoption in Bihar, Odisha and Madhya Pradesh i.e., videos with similar content to the previously adopted ones have a higher chance of adoption. Seventh, the active age of association of a farmer with DG negatively impacts their adoption across all the states. We plan to investigate the underlying reasons for the same in our future work. Eighth, farmers in Bihar, Andhra Pradesh and Madhya Pradesh benefit from higher mediator attention (MA) i.e., videos attended by fewer farmers in one screening are preferred. Lastly, while the timing at which the video is screened plays a relatively less important role in influencing adoption, the SHAP plots give an overview of the varying timing preferences of farmers across the states.

[13] For Content Specificity (CS), V denotes village, B denotes block and D denotes district.

Table 6: Classification Results with macro-f1 and TN Rate for the three models across the five states. Random Forest produces the best results for both metrics.

	Logistic Reg		XGBoost		Random Forest	
State	Macro-F1	TN	Macro-F1	TN	Macro-F1	TN
Bihar	0.61	0.59	0.65	0.63	**0.89**	**0.90**
Andhra Pradesh	0.62	0.66	0.65	0.63	**0.85**	**0.84**
Odisha	0.72	0.71	0.75	0.71	**0.87**	**0.85**
Madhya Pradesh	0.57	0.56	0.61	0.59	**0.79**	**0.79**
Karnataka	0.64	0.62	0.71	0.70	**0.82**	**0.82**

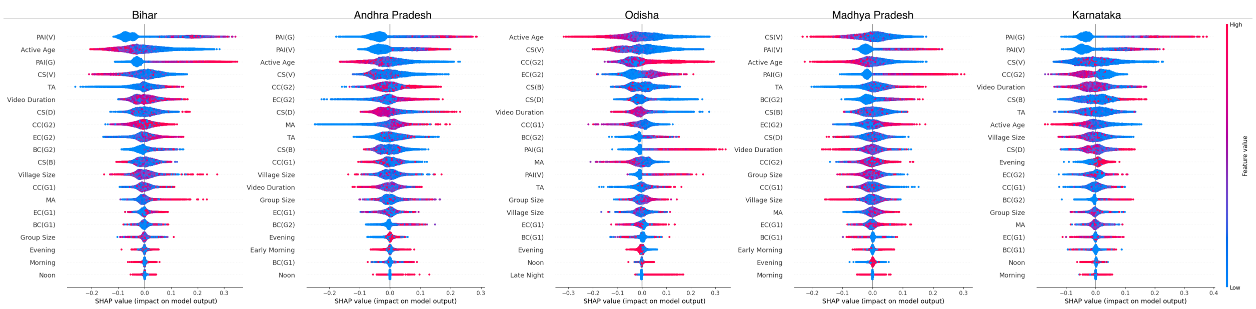


Figure 7: SHAP summary plots for model features for all five states, with features ranked by importance.

6 Implications

In this section, we highlight the implications of our findings from Sections 4 and 5.2 for Digital Green with considerations for implementing them.

i) Focused Assistance: Our model accurately identifies farmers who might face challenges in adopting videos. This will enable Digital Green to assist them with training and resources or conduct ethnographic investigation to better understand their difficulties.

ii) Community Building: Both statistical tests and model outcomes determine that past adopters of a video can greatly help farmers in their local community in adopting the video. Hence, we suggest building co-adoption communities at the ground level for farmers with low PAI to alleviate their adoption rates. In cases where significant farmers in a village/group have not adopted a video and their past co-adopters at the group-village levels have adopted it, a recommendation to screen that video can also be made.

iii) Recommendations for Videos: In Bihar, Odisha and Madhya Pradesh, we suggest the use of titles for videos that have high TA due to its positive impact on adoptions. In cases where the videos cannot be represented by such titles due to difference in keywords, we recommend revising the video content so that it fits in a shorter duration across Odisha and Madhya Pradesh. We suggest the same for videos in Andhra Pradesh and Karnataka because of the negative impact of longer videos on adoptions. Further, both statistical tests and model explanations indicate that higher mediator attention positively impacts adoptions in Bihar, Andhra Pradesh and Madhya Pradesh. Hence, we recommend improving the farmer:mediator ratio for farmers with lower adoption rates.

iv) Rethinking Participatory Approach: In the previous section, we found that the content specificity had a mostly negative impact on adoptions as per our model which was contrary to our understanding. This indicates that their participatory content production process might require further diagnosis on ground. We suggest the same and prescribe making their participatory approach more inclusive by ensuring representation across multiple axes of marginalisation including caste, class and gender.

v) Mitigating Inequalities: In Section 4.2, we found significant gender-based inequalities in terms of adoption rates across three states. This will enable Digital Green in investigating and mitigating them as their service continues.

7 Discussion

In this work, we looked at ten years of data from the web-based data tracker of an ICT (Digital Green) that seeks to empower rural households by enabling knowledge sharing of various types of practices. In particular, we examined the adoption of agricultural practices across five states of India – Bihar, Andhra Pradesh, Madhya Pradesh and Karnataka. We modelled different components of the Digital Green ecosystem and used statistical methods to identify various factors that distinguish farmers with higher adoption rates from others. We diagnosed the Digital Green ecosystem to highlight gender-based inequalities among farmers in Andhra Pradesh, Odisha and Karnataka. While our analysis is currently limited to gender, we plan to include class and caste in our future work to investigate inequalities from an intersectional feminist lens. Next, we leveraged the modelled features and experimented with different classifiers to accurately identify farmers who might face challenges in adopting videos. We argue that this would further enable us to conduct fieldwork and ethnographic inquiry into their experiences, allowing us to account for how power dynamics unfold locally. We explained our model results using SHAP plots and verified some of our claims by delving deeper into the dependency plots in Appendix Section B. Lastly, we aggregated our findings to provide implications for alleviating adoption rates of nearly a million farmers in the Digital Green ecosystem. Our research builds upon past literature by studying farmer network dynamics and their role in adoptions of diverse farming practices at scale. It also sheds light on key factors for effective information dissemination such as video characteristics and mediator-farmer ratio.

We acknowledge that our findings are quantitative and serve as a diagnosis for the ICT. As a result, our implications rely on qualitative fieldwork to generate experiential considerations before implementation. To this end, we plan to use a mixed-methods approach for our future work to account for on-field experiences of farmers. We will also evaluate the effectiveness of self help groups to explore whether a bottom-up approach might be more beneficial as compared to a top-down policy driven approach by the government. Finally, we aspire to expand our research to non-agricultural practices and the remaining seven states in India where Digital Green is operational.

References

[1] India Brand Equity Foundation. Analysis about indian agriculture industry, market size, export & investment opportunity, Jul 2017.

[2] Rikin Gandhi, Rajesh Veeraraghavan, Kentaro Toyama, and Vanaja Ramprasad. Digital green: Participatory video for agricultural extension. In *2007 International Conference on Information and Communication Technologies and Development*, pages 1–10, 2007.

[3] Richard Heeks. The ict4d 2.0 manifesto: Where next for icts and international development? *Development Informatics Working Paper*, (42), 2009.

[4] S. Patel, O. Bataveljic, P.J. Lisboa, C. Hawkins, and R. Rajan. ishakti–crossing the digital divide in rural india. In *2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)*, pages 1061–1065, 2006.

[5] Divya Ramachandran, John Canny, Prabhu Dutta Das, and Edward Cutrell. *Mobile-Izing Health Workers in Rural India*, page 1889–1898. Association for Computing Machinery, New York, NY, USA, 2010.

[6] Nicola Dell and Neha Kumar. The ins and outs of hci for development. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 2220–2232, New York, NY, USA, 2016. Association for Computing Machinery.

[7] Nithya Sambasivan and Thomas Smyth. The human infrastructure of ictd. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, ICTD '10, New York, NY, USA, 2010. Association for Computing Machinery.

[8] Sebastien Cuendet, Indrani Medhi, Kalika Bali, and Edward Cutrell. *VideoKheti: Making Video Content Accessible to Low-Literate and Novice Users*, page 2833–2842. Association for Computing Machinery, New York, NY, USA, 2013.

[9] Kalika Bali, Sunayana Sitaram, Sebastien Cuendet, and Indrani Medhi. A hindi speech recognizer for an agricultural video search application. In *Proceedings of the 3rd ACM Symposium on Computing for Development*, ACM DEV '13, New York, NY, USA, 2013. Association for Computing Machinery.

[10] S. Mittal and M. Mehar. *Agricultural information networks, information needs and risk management strategies: A survey of farmers in Indo-Gangetic Plains of India*. Socioeconomics Working Paper. CIMMYT.

[11] Gershon Feder and Dina L. Umali. The adoption of agricultural innovations: A review. *Technological Forecasting and Social Change*, 43(3):215–239, 1993. Special Issue Technology and Innovation In Agriculture and Natural Resources.

[12] Duncan Knowler and Ben Bradshaw. Farmers' adoption of conservation agriculture: A review and synthesis of recent research. *Food Policy*, 32(1):25–48, 2007.

[13] Margaret R Mwangi and Sam Kariuki. Factors determining adoption of new agricultural technology by smallholder farmers in developing countries. *Journal of economics and sustainable development*, 6:208–216, 2015.

[14] Andreas Kamilaris, Andreas Kartakoullis, and Francesc X. Prenafeta-Boldú. A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, 143:23–37, 2017.

[15] R. Hinz, T. B. Sulser, R. Huefner, D. Mason-D'Croz, S. Dunston, S. Nautiyal, C. Ringler, J. Schuengel, P. Tikhile, F. Wimmer, and R. Schaldach. Agricultural development and land use change in india: A scenario analysis of trade-offs between un sustainable development goals (sdgs). *Earth's Future*, 8(2), 2020.

[16] K. W. Maina, C. N. Ritho, B. A. Lukuyu, and E. J. O. Rao. Socio-economic determinants and impact of adopting climate-smart brachiaria grass among dairy farmers in eastern and western regions of kenya. *Heliyon*, 6(6):e04335–e04335, Jun 2020. 32637709[pmid].

[17] Surabhi Mittal and Mamta Mehar. Socio-economic factors affecting adoption of modern information and communication technology by farmers in india: Analysis using multivariate probit model. *The Journal of Agricultural Education and Extension*, 22(2):199–212, 2016.

[18] Minal Sawant, Rajesh Urkude, and Sandip Jawale. Organized data and information for efficacious agriculture using pride™ model. *The International Food and Agribusiness Management Review*, 19:115–130, 2016.

[19] Jenny C. Aker, Ishita Ghosh, and Jenna Burrell. The promise (and pitfalls) of ict for agriculture initiatives. *Agricultural Economics*, 47(S1):35–48, 2016.

[20] Nidhi Tewathia, Anant Kamath, and P. Vigneswara Ilavarasan. Social inequalities, fundamental inequities, and recurring of the digital divide: Insights from india. *Technology in Society*, 61:101251, 2020.

[21] Tian Lin, Aung Phyo Ko, Maung Maung Than, Delia C. Catacutan, Robert F. Finlayson, and Marney E. Isaac. Farmer social networks: The role of advice ties and organizational leadership in agroforestry adoption. *PLOS ONE*, 16(8):1–18, 08 2021.

[22] Timothy Conley and Christopher Udry. Social learning through networks: The adoption of new agricultural technologies in ghana. *American Journal of Agricultural Economics*, 83(3):668–673, 2001.

[23] Annemie Maertens and Christopher B. Barrett. Measuring social networks' effects on agricultural technology adoption. *American Journal of Agricultural Economics*, 95(2):353–359, 2013.

[24] D. J. Pannell, G. R. Marshall, N. Barr, A. Curtis, F. Vanclay, and R. Wilkinson. Understanding and promoting adoption of conservation practices by rural landholders. *Australian Journal of Experimental Agriculture*, 46(11):1407–1424, 2006.

[25] Francisco Alcon, Sorada Tapsuwan, José M. Martínez-Paz, Roy Brouwer, and María D. de Miguel. Forecasting deficit irrigation adoption using a mixed stakeholder assessment methodology. *Technological Forecasting and Social Change*, 83:183–193, 2014.

[26] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 4768–4777, Red Hook, NY, USA, 2017. Curran Associates Inc.

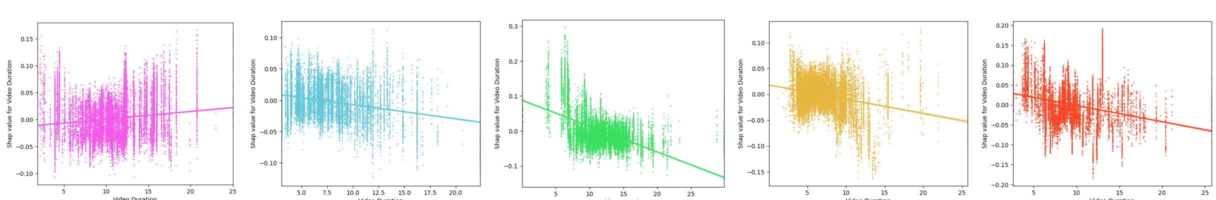
[27] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery.

A Mean Values for Table 3

Mean values for both q_1 and q_4 across all the eight factors corresponding to the One-Tailed Welch's t-test are given in Table 7, grouped on the basis of hypotheses.

B Shapley Dependency Plots

This section contains various SHAP Dependency Plots to capture the relationship between certain features and their impact on the model output in detail. Figure 8 represents the SHAP dependency plot for video duration. Figures 9, 10 and 11 represent the SHAP dependency plots for content specificity at the village, block and district levels respectively.



(a) Bihar (b) Andhra Pradesh (c) Odisha (d) Madhya Pradesh (e) Karnataka

Figure 8: SHAP dependency plots for video duration across the five states when fit with a regression line. Only Bihar has a positive slope, i.e., videos of longer duration have a positive impact on adoption.

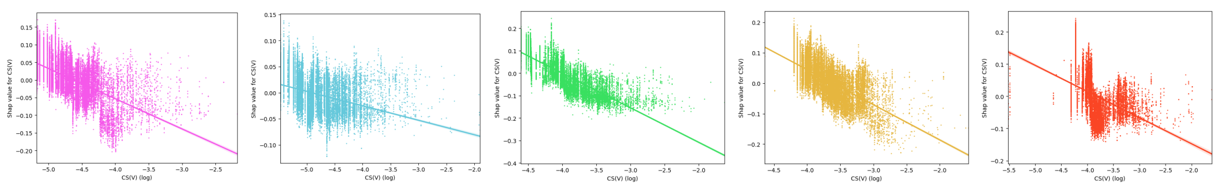
Table 7: Mean Values corresponding to One-Tailed Welch t-Tests for eight factors between farmers in q_1 and q_4

(a) $H_1 : q_1 < q_4$

State	MA_μ		$CS_{V\mu}$		$PAI_{G\mu}$		$PAI_{V\mu}$		$Active\ Age$	
	q_1	q_4	q_1	q_4	q_1	q_4	q_1	q_4	q_1	q_4
Bihar	0.048	0.052	$4.78e^{-5}$	$3.73e^{-5}$	0.448	0.519	8.21	13.41	637.53	191.52
Andhra Pradesh	0.060	0.066	$3.49e^{-5}$	$6.58e^{-5}$	4.33	2.56	17.58	13.45	266.26	57.66
Odisha	0.061	0.046	$1.93e^{-4}$	$9.31e^{-5}$	0.279	0.690	2.87	5.32	915.81	276.22
Madhya Pradesh	0.069	0.077	$3.41e^{-4}$	$3.41e^{-4}$	0.396	0.625	2.29	3.31	506.48	323.43
Karnataka	0.099	0.090	$2.18e^{-4}$	$7.75e^{-4}$	0.387	0.587	4.81	5.83	664.18	178.07

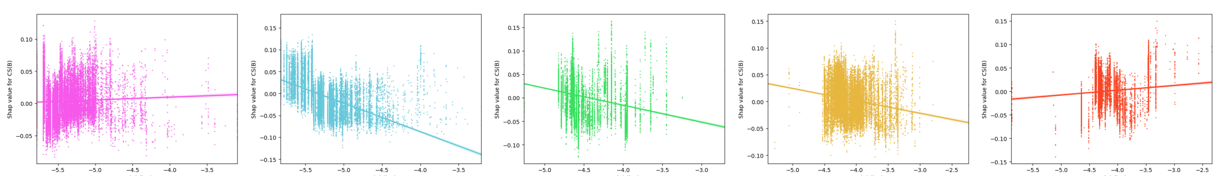
(b) $H_1 : q_4 < q_1$

State	$duration_\mu$		GS		VS	
	q_1	q_4	q_1	q_4	q_1	q_4
Bihar	10.57	10.39	12.20	11.55	156.62	147.79
Andhra Pradesh	7.61	7.53	22.41	23.09	185.27	159.61
Odisha	12.14	10.49	16.69	21.04	163.88	118.91
Madhya Pradesh	7.38	7.16	17.70	15.63	98.82	86.31
Karnataka	9.76	8.75	13.52	15.87	247.25	132.07



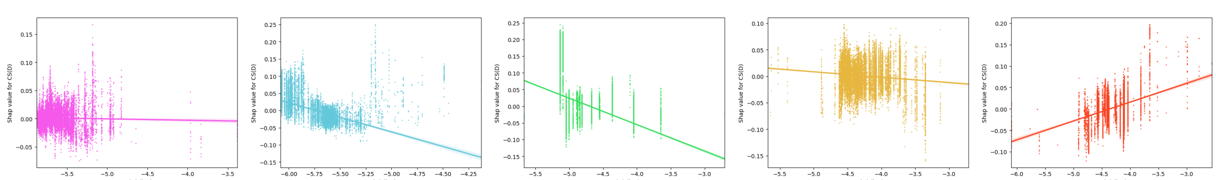
(a) Bihar (b) Andhra Pradesh (c) Odisha (d) Madhya Pradesh (e) Karnataka

Figure 9: SHAP dependency plots for CS_V across the five states when fit with a regression line. It has a negative impact on adoption across all the states.



(a) Bihar (b) Andhra Pradesh (c) Odisha (d) Madhya Pradesh (e) Karnataka

Figure 10: SHAP dependency plots for CS_B across the five states when fit with a regression line. We see a negative impact on adoption across all states except Bihar and Karnataka.



(a) Bihar (b) Andhra Pradesh (c) Odisha (d) Madhya Pradesh (e) Karnataka

Figure 11: SHAP dependency plots for CS_D across the five states when fit with a regression line. We only see a positive impact on adoptions in Karnataka.