

Received July 13, 2021, accepted July 28, 2021, date of publication August 6, 2021, date of current version August 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3103046

# Human-Survey Interaction (HSI): A Study on Integrity of Human Data Collectors in a Mass-Scale Hajj Pilgrimage Survey

**SAIFUL ISLAM SALIM<sup>1,\*</sup>, NAJLA ABDULRAHMAN AL-NABHAN<sup>2,\*</sup>, MASFIQUR RAHAMAN<sup>1</sup>, NAFISA ISLAM<sup>3,4</sup>, TARIK REZA TOHA<sup>1,5</sup>, JANNATUN NOOR<sup>1,6</sup>, ADNAN QUAJUM<sup>1,7</sup>, AAIYEESHA MOSTAK<sup>1</sup>, MAINUL HOSSAIN<sup>1</sup>, MD. MASUM MUSHFIQ<sup>1</sup>, AND A. B. M. ALIM AL ISLAM<sup>1</sup>**

<sup>1</sup>Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh

<sup>2</sup>Department of Computer Science, King Saud University, Riyadh 11451, Saudi Arabia

<sup>3</sup>Department of Civil Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh

<sup>4</sup>Department of Civil and Environmental Engineering, Islamic University of Technology, Gazipur 1704, Bangladesh

<sup>5</sup>Department of Textile Machinery Design and Maintenance, Bangladeshi University of Textiles, Dhaka 1208, Bangladesh

<sup>6</sup>Department of Computer Science and Engineering, BRAC University, Dhaka 1212, Bangladesh

<sup>7</sup>Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, Dhaka 1208, Bangladesh

\*Saiful Islam Salim and Najla Abdulrahman Al-Nabhan contributed equally to this work.

Corresponding author: Masfiqur Rahaman (1505111.mr@ugrad.cse.buet.ac.bd)

This work was supported by the Deputyship for Research and Innovation, Ministry of Education, Saudi Arabia, under Project DRI-KSU-762.

**ABSTRACT** Mass gatherings (such as Hajj/Umrab), owing to their immensity, often present a variety of difficulties to the attendees. To have a comprehensive understanding of the difficulties as well as their potential remedies, surveying a good number of attendees is unavoidable and this can be facilitated through engaging data collectors. A crucial part here is identifying the integrity of the data collectors, which is yet to be explored in the literature to the best of our knowledge. To address this gap, in this study, we first perform a mass-scale data collection over Hajj/Umrab pilgrims through online ( $n = 236$ ) and in-person ( $n = 752$ ) surveys, where we cover a substantial part ( $n = 712$ ) through paid data collectors ( $n = 53$ ). We critically investigate data collection activities of the data collectors through focused group discussions involving expert reviewers. We explore two computing techniques to unveil the integrity of the data collectors from two different perspectives. Our study finds out influential (religious and socio-geographical) aspects that impact the data collection process. Besides, we find that the collaborative participation of expert reviewers is obligatory to scrutinize the data collectors' integrity. Additionally, our explored computing techniques, namely conflict analysis and learning-based analysis, can identify up to 74% and 99% of the unreliable data collectors respectively. We observe that, although these computing-based filtering can indicate the integrity up to a certain level, human-in-the-loop is unavoidable for concluding on the integrity. To the best of our knowledge, this is the first study of its kind (i.e., integrity analysis of the data collectors) in the literature.

**INDEX TERMS** Collaborative filtering, religious congregation, social computing, empirical investigations, data collector integrity, reliability.

## I. INTRODUCTION

Exploring collaboration between researchers and online volunteers is one of the prominent fields of research in social computing [1]–[9]. Volunteering is increasingly recognized as a significant aspect of collaborative system design, ranging

The associate editor coordinating the review of this manuscript and approving it for publication was Saqib Saeed<sup>1</sup>.

from a survey study to online crowd-sourcing. Such participation of volunteers is inevitable in the case of mass-scale survey studies, which enables covering a significant number of participants to obtain a thorough understanding of a topic of interest. An example in this regard is gathering information on the experiences of pilgrims during Hajj, which is one of the major annual mass religious congregations all over the globe. To develop a comprehensive idea of the experiences of

pilgrims, a mass-scale survey needs to be performed where engagement of a significant number of volunteers or data collectors is required to collect data from pilgrims. Pilgrims attend different religious congregations worldwide having millions of attendees gathering even in a single event [10]. Examples in this regard include the Hajj in Mecca of Saudi Arabia, Kumbh Mela in Prayagraj of India, Arba'een pilgrimage in Karbala of Iraq, pilgrimage in Lourdes of France, etc. Among them, the annual Hajj of Mecca in Saudi Arabia is the best-known religious mass gathering annual event having the participation of 2.5 million pilgrims each year [11].

The event of Hajj is considered to be the fifth pillar of the Islamic faith [12] and is obligatory for all well-off Muslims to perform Hajj at least once in their lifetime. It is the biggest annual conglomeration of people on earth, where people from more than 140 countries perform several rituals for three to five days between the eighth and twelfth of Dhul-Hijjah, the last month of the Islamic calendar [12]. During Hajj, major rituals include circulating Kaa'ba in Mecca, stoning at Jamrats and slaughtering animals in Mina, gathering in Arafah, spending a particular night in Muzdalifah, etc. All of the rituals are associated with huge mass gatherings and heavy crowds. Due to overcrowding, pilgrims experience numerous problems including health issues, sickness, heat strokes, unavailability of food and water, the inadequacy of washroom, pick-pocketing and theft, losing necessary documents, getting lost, communication difficulties, etc. Besides, many mishaps also occurred in previous years, causing people to be killed by stampede and fire, crushed by the failure of construction cranes, trampled during stoning rituals, etc. For instance, in 1990, around 1426 pilgrims were killed during a stampede inside a pedestrian tunnel (Al-Ma'asim tunnel) [13], [14]. In 1997, around 343 pilgrims were killed due to a tent fire in Mina [14]. In 2006, a total of 346 pilgrims were killed during stoning in Mina [13]–[15]. In 2015, around 2431 pilgrims were killed and 427 pilgrims from 42 countries were also reported missing during a stampede in Mina after overcrowding of huge masses in the blocked street of Mina leading up to Jamaraat [16], [17]. In the same year, in 2015, a crush outside Mecca killed 717 pilgrims and injured at least 863 pilgrims [18].

Accordingly, the difficulties and problems faced by pilgrims are diversified as well as the level of severity and commonality of the problems are different. As such, to obtain a comprehensive understanding of pilgrims' experiences during Hajj, the conduction of a mass-scale survey over the pilgrims' is unavoidable.

To conduct such a mass-scale survey, engaging *data collectors*<sup>1</sup> is a primary requirement [19]. Since such a

<sup>1</sup>We conduct a mass-scale survey through both in-person and online data collection, and we engage personnel from different institutions (see Section III-D) as *data collectors* for successfully conducting the in-person data collection. We offer remuneration to the personnel for their data collection work. Note that, as the personnel often need to spend substantial time and effort only for the purpose of data collection, we prefer to use the term *data collector* for them rather than *volunteer/surveyor*.

survey requires responses from pilgrims having different socio-cultural backgrounds, engaging data collectors from different places and diverse contexts exhibit a direct necessity. As the locations and contexts of the data collectors need to vary, their engagement through an online system would be a suitable approach to ensure reaching as well as maintaining communication with this diversified group of people.

It is imperative that, as we grow the engagement, so the possibility of getting more survey data gets increased. With the increase in the possibility of getting more survey data, the concern on the quality of collected survey data also comes to light. The quality of collected survey data in turn depends on the integrity of the data collectors. To confirm the integrity of the data collectors, their activities and quality of their collected data need to be thoroughly assessed.

To ensure the quality of collected survey data, existing research studies explored ways of improving performances of data contributors or survey participants as the data collection process progresses [1], [4], [20], [21]. Besides, the focus of other related studies was on improving the engagement of survey participants who directly contribute to the data collection through providing their own data [1], [4]. These studies do not focus on engaging any data collector in the process of data collection. Nonetheless, despite the progress in assessing the roles of computing techniques in survey data validation, the contribution of computing techniques to assess the integrity of data collectors is yet to be explored. In this regard, the collaborative process exhibited a noteworthy role, however, only in the process of survey data validation [1]. Besides, the applicability of such collaborative processes has not been investigated in the case of determining the integrity of data collectors. In summary, existing research studies are yet to focus on assessing the integrity of data collectors.

Therefore, in this study, we specifically focus on the integrity of data collectors, who perform in-person surveys with the survey participants and collect data from them. Here, we analyze different activities performed by the data collectors as well as investigate applicable scrutinizing aspects (both computing technique-based and manual) for them from the context of a mass-scale survey. In the process of the study, we adopt two different modalities of survey data collection - in-person and online - for surveying experiences of Hajj/Umrab pilgrims. In the case of the online survey, we separately perform surveys in the native (Bangla) language covering only natives (Bangladeshi) and in the English language covering both natives and other nationals. In the case of an in-person survey, we separately perform surveys through the authors and paid data collectors. Here, we onboard the data collectors to collect large-scale pilgrim survey data. We methodically perform data collector recruitment, provide training, supervise the data collection process, provide their remuneration, and finally get feedback from them.

We engage the data collectors in the processes of training and supervising through online platforms. After being trained, the data collectors perform in-person surveys with the pilgrims and record their experiences. Throughout the

data collection process, we examine the activities of the data collectors. Besides, we adopt an online collaborative scrutinizing process as well as a computing technique to evaluate the integrity and reliability of the data collectors. Here, integrity refers to collecting responses that are not maliciously fabricated. Besides, reliability refers to collecting responses while being aligned with applicable terms and conditions of data collection, which is challenging to be due to socio-geographical aspects of the study location as described later. While scrutinizing the activities of data collectors during the data collection process, we consider several factors to classify a data collector as reliable or unreliable. The influential factors include the degree of alignment with our survey protocol, outcomes of aggregated analysis, conversation with both participants and data collectors, data collector's response rate, data collector's effort, and participants' religious contexts. We consider these factors to categorize the data collectors during our focused group discussion. Thus, in the process of the whole study, we aim to answer the following research questions.

#### A. RESEARCH QUESTIONS

Based on the above contexts, we explore the following research questions in this study.

*RQ1: What are the activities performed by human data collectors during mass-scale survey data collection that affects the quality of collected Hajj pilgrimage survey data? How do the activities affect the quality of the survey data?*

In the case of engaging human data collectors for recording experiences of participants in a mass-scale survey, it is likely to find diverse activities performed by the data collectors during the data collection process. These activities of data collectors can greatly influence the quality of the collected data. Besides, reactions exerted by the data collectors throughout the data collection process can also vary. Moreover, these reactions can exhibit an influence on the quality of the collected data. We aim to analyze the influences on the quality of the collected data in this paper.

*RQ2: What are the factors that influence the collaborative scrutinizing of data collectors in a mass-scale Hajj pilgrimage survey?*

The integrity of data collectors, who collect data by surveying participants, can impact the quality of collected data. Therefore, evaluation of the integrity of data collectors is necessary to ensure the credibility of the collected data. Collaborative scrutiny by expert reviewers, as explored in different studies in the literature [1]–[3], could be highly effective in this regard. However, it is unclear that, to what extent, collaborative scrutiny can help in evaluating the integrity of data collectors. Besides, it is not evident how different factors such as response rate, alignment with survey protocol, etc., influence the task of collaborative scrutiny. Therefore, we aim to analyze the influences of different factors on the collaborative scrutinizing of data collectors in this paper.

*RQ3: Can conflict analysis measure the integrity of data collectors in a mass-scale Hajj pilgrimage survey?*

Conflict analysis represents discrepancies in the survey response of a particular pilgrim. Deceitful and indifferent collection of data by the data collectors might result in inconsistency and conflict in their collected responses. These conflicts can indicate effective identifications of possible unreliable data collectors. However, it is not evident how far these conflicts portray the integrity of the data collectors. Therefore, we aim to study to what extent conflict analysis can measure the integrity of data collectors.

*RQ4: Can standard machine learning distinguish between reliable and unreliable data collectors in a mass-scale Hajj pilgrimage survey?*

Adoption of computing techniques can greatly assist in the process of scrutinizing data collectors to distinguish between reliable and unreliable ones. Here, standard machine learning could be effective to identify and flag possible unreliable data collectors. We study to what extent machine learning can benefit the identification of potentially unreliable data collectors. This study eventually portrays the capability and limitation of computing techniques in the process of measuring reliability and integrity of data collectors, and thus, identify potential human-in-the-loop in the process.

#### B. OUR CONTRIBUTIONS

To resolve the above research questions, we make several contributions in this paper as follows.

- First of all, we perform a mass-scale data collection for identifying different types of difficulties experienced by the pilgrims while performing Hajj/Umrah. We conduct our data collection through in-person ( $n = 752$ ) and online ( $n = 236$ ) surveys. Among the participants surveyed in-person, a substantial part ( $n = 712$ ) were surveyed through paid data collectors.
- We closely examine the activities of the paid data collectors and check for clarification in case of detecting any anomaly. Our conversations with the data collectors to clarify the detected anomalies reveal diversified activities as well as reactions of individual paid data collectors. Accordingly, we explore the pattern of events of falsification exercised throughout the data collection process by the data collectors. We also analyze the activities performed by the data collectors in response to different religious and socio-geographical aspects experienced during the data collection process. Our study reveals that religious and socio-geographical aspects exhibit both favorable and unfavorable impacts on the data collection process of a mass-scale survey.
- Besides, we perform rigorous collaborative scrutiny over the collected data through focused group discussion during the data collection process. Through the scrutiny, we observe that there exist several aspects that exhibit significant influence over the integrity of data collectors. Examples include the response rates, presence of a conflict in collected information, degree of alignment during the data collection process, socio-geographical

aspects, etc. Subsequently, based on our findings we propose a pathway for effective collaborative scrutiny over the data collectors.

- In addition, we investigate effective assessment of the integrity of data collectors. Here, we measure the integrity of data collectors through employing information-based conflict analysis. Our conflict analysis can identify 74% of the unreliable data collectors.
- Finally, we investigate the effectiveness of k-means clustering to detect unreliable data collectors. Leveraging the clustering-based approach, we can identify 99% of the unreliable data collectors.

## II. RELATED WORK

To get a useful insight regarding the diversified experiences pilgrims face during Hajj, we need to perform a large-scale survey. However, only a few existing research studies performed such surveys on the event of Hajj. For example, the study in [22] performed a survey on 154 pilgrims and examined the paradigm of human psychology during Umrah in Mecca. Another study surveyed 1,176 pilgrims focusing on psychological domain [23]. They observed that traits of disregard for their own safety, cultural differences in individual reception and response, etc., might contribute to casualty cases. To warn pilgrims about these casualty cases, a recent study [12] investigated the efficiency of different information sources (billboards, TV news channels, brochures, etc.) through a survey of over 348 pilgrims. Besides, another study conducted an online survey of over 215 pilgrims to explore the adequacy of existing road sign languages for pilgrims and proposed an improved system for better visibility [24]. In addition, to explore the degree of potential and usefulness of various technologies such as wireless networking, computer vision, spatial computing, etc., in terms of crowd management during Hajj, the study in [25] conducted a survey. However, these studies did not focus on identifying a comprehensive view on diversified difficulties pilgrims faced during Hajj. Instead, each of these studies addressed a single event of interest. However, to improve the overall experiences of pilgrims during Hajj, a comprehensive study should be performed to identify all possible difficulties that pilgrims might face during Hajj. A large-scale survey study, by engaging human resources or data collectors who will collect data by surveying over the Hajj/Umrah pilgrims, is envisioned to help in this regard.

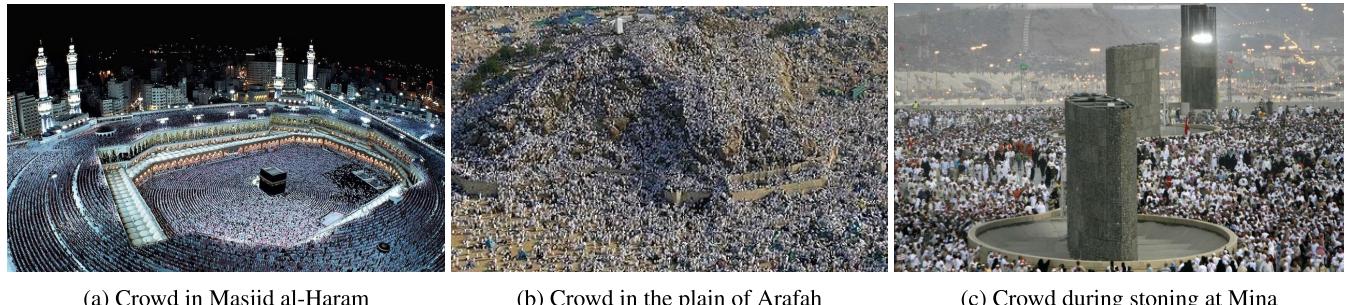
Considering the aspects discussed above, we explore the existing literature from three different research perspectives - 1) research on human resource, i.e., data collectors engagement in survey work, 2) research on psychological aspects of engaged data collectors that can influence outcomes of the data collection, and 3) research on technological artifacts that can facilitate evaluating the integrity of engaged data collectors. Therefore, in the following, we delineate related research studies from the existing literature covering all these three different perspectives.

1) HUMAN RESOURCE ENGAGEMENT IN SURVEY RESEARCH  
Performing mass-scale community work needs human resources such as data contributors, data collectors, etc. Data contributors can be engaged when they are required to contribute directly to a study by providing their own data. Here, data collectors can be engaged as intermediaries who survey the data contributors and record the data contributors' data. Engaging data collectors in a survey study can save time and cover a wide range of demographic and geographic variations. In addition to these specific contexts, numerous research fields get significant contributions through engaging data contributors and data collectors [1], [4]–[6], [26].

Besides, several research studies aimed at developing an understanding of the infrastructure and the ways utilized by data contributors [4], [7], [21], [27]. In this regard, the study in [7] provided a comprehensive idea on the social work structure of data contributors and investigated the roles of technology that helps in performing the data contribution activities. In addition, the study in [27] adopted the theory of Activity Analysis to understand cooperative work structure in a hospital environment context. Here, instead of taking an ethnmethodological approach, they proved Activity Theory to be useful in performing the study, analyze, describe, and understand human activities covering collaborative activities. Further, the study in [4] proposed three analytical approaches to assess contributor engagement: cohort analysis, task analysis, and observation over contributor performance. They stated that coordination practices, task designs, individual contributor characteristics, etc., have a marked impact on contributor performance.

On the other hand, the study in [21] proposed a technique to determine the expected retention period of new contributors for online projects based on their first session. The retention period is very important in successfully engaging data contributions in a sustainable way. They introduced a metric named 'DSCORE' that can anticipate new contributor longevity based on their early activities. They showed that 'DSCORE' is useful in case of measuring diversity over new contributors' activities. Nonetheless, an incentive is also noticed to have a profound impact on engaging data contributors. Several studies examined the role of incentives on the participation of data contributors [28], [29].

All these studies mostly focused on improving the engagement and performance of data contributors. However, in the case of engaging data collectors in the survey process who survey data contributors to collect responses, it is yet to be studied how the activities of data collectors influence the progression of the survey. Though understanding the data collectors' dynamics is also important for achieving successful outcomes from a survey, this aspect is also little explored in the literature. Besides, for the purposes of analyzing activities and dynamics of data collectors, a framework for collaborative scrutinizing over the data collectors is needed. However, these aspects pertinent to data collectors are yet to be investigated in the literature, to the best of our knowledge.



(a) Crowd in Masjid al-Haram

(b) Crowd in the plain of Arafah

(c) Crowd during stoning at Mina

**FIGURE 1.** Crowd generation during different places and events of Hajj.

## 2) PSYCHOLOGICAL ASPECTS OF ENGAGED HUMAN RESOURCE IN SURVEY WORK

Along with improving performance and increasing the longevity of the engaged human resources in a survey, their integrity, trust, and reliability to the survey need to be ensured. In this regard, research studies focused on the issue of trust among engaged human resources and identified the trust issue as a major contributor to the outcomes of cooperative work such as survey, be it a success or failure [30]–[32]. Hence, the study in [30] presented a set of design patterns that aim at fostering trust among members of a computer-supported collaborative work of highway maintenance. They developed these design patterns through their own learning process (development, testing, and refinement) and showed that these patterns can efficiently delineate the contribution of trust in dynamic cooperative work.

The impact of trust in collaborative work, in turn, depends on several factors. The study in [32] pointed out the importance of “baseline trust” (a personal concept of trust in each individual) on the development and evolution of trust among team members in Global Software Engineering (GSE) practices. They designed a virtual experiment that integrates agent-based modeling with a realistic network structure and showed that outside-work light talks help improving trust and cooperation among virtual team members. Besides, project monitoring is observed to exhibit a significant impact on trust development in cooperative work structures. In this regard, the work in [31] studied 57 virtual teams and examined the impacts of two types of monitoring (internal monitoring and external monitoring) on two types of trust (affective trust and cognitive trust) development. Here, monitoring by team members is termed as internal monitoring, and monitoring by an outside supervisor is termed as external monitoring. The study observed that, under internal and external monitoring, affective trust increases and decreases respectively. On the other hand, under all types of monitoring, cognitive trust decreases. Thus, these existing studies focused on observing and improving the performance of engaged human resources in cooperative work. Their findings indicate that the individual mindset of contributors, their sincerity to the work, effective coordination, and monitoring, etc., play a significant role in mediating the progress of the work. However, in the

case of engaging human resources for data collection in a large-scale survey, several important psychological aspects such as to what extent trust and integrity of the data collectors can be achieved, what are types of falsifications the data collectors can adopt, etc., still remain under-explored.

## 3) EXISTING COMPUTING METHODS FOR SCRUTINIZING ENGAGED HUMAN RESOURCES

Evaluating the integrity of data collectors is critical in the road to ensuring the credibility of data collected by them. To assess their integrity, in addition to understanding their psychology, exploring the ways through which different computing methods can facilitate the evaluation process is also important. Here, the introduction of computing methods can ease the process of identifying human resources having integrity. However, this field of research i.e., exploring computing methods in the process of evaluating the integrity of human resources engaged in a survey process, is little investigated in the literature. Few research studies in this regard explored the efficacy of different computing techniques to measure the performance of data collectors. For example, the study in [4] evaluated first-time contributor engagement using cohort and task analyses. Under the cohort analysis, they used Spearman Correlation analysis, Kolmogorov Pairwise analysis, and Kaplan-Meier Estimates to determine retention rates of the data contributors, degree of their prior experience, and their survival rates, etc. Under the task analysis, they prepared a regression model to determine early abandonment based on early user experience and the size of the work. Besides, the study in [33] explored the ways of conflict management in online collaboration work and adopted the Cox Proportional-Hazards regression model for survival analysis over the associated human resources. Additionally, another study [20] used random-effects regression models and negative binomial regression models to measure retention periods of individual newcomers and their early activities. In addition to investigation on performances of the data contributors, approaches have been studied to validate the data contributed by the data contributors. Here, the study in [1] observed some collaborative interaction patterns performed by expert reviewers and compared the efficacy of these patterns for improvement of collected data.

**TABLE 1.** Summary of the survey questionnaire.

Section no.	Section	Question no.	Response format
1	Demographic information	1-12	Simple text, binary-choice, and multiple-choice
2	Communication difficulties	13-14	Binary-choice
3	Information about getting lost	15-23	Binary-choice and multiple-choice
4	Vehicle shortage	24-25	Binary-choice and multiple-choice
5	Washroom shortage	26-29	Binary-choice and multiple-choice
6	Drinking water shortage	30-31	Binary-choice and multiple-choice
7	Food shortage	32-33	Binary-choice and multiple-choice
8	Problems in Masjid al-Haram	34-37	Binary-choice and multiple-choice
9	Problems in Mina camp	38-40	Binary-choice and multiple-choice
10	Problems in 'Arafah	41-43	Binary-choice and multiple-choice
11	Road directions in Hajj-related places	44-47	Binary-choice and multiple-choice
12	Miscellaneous information	48-49	Multiple-choice

Apart from the collaborative interaction and other methods discussed above, different statistical and machine learning methods are becoming popular in recent studies [34]–[36]. However, those studies specifically focus on the quality of data and the data warehouse where the data is managed efficiently. As data model influences data warehouse quality, the study in [34] used statistical and machine learning methods to predict the effect of structural metrics on the efficiency and effectiveness of a conceptual model. Nonetheless, the study in [35] presented a tool for continuously monitoring the quality of data to increase the prediction accuracy of machine learning models.

All these existing research studies focused on improving the quality of data warehouses or the data itself. However, they do not cover the aspect of evaluating the integrity of data collectors, when they are engaged in large-scale survey work. Nonetheless, how a collaborative scrutinizing process with the help of computing methods can facilitate coordinators to identify data collectors having integrity is yet to be examined. Therefore, despite increased research on survey studies and data contributor management, to the best of our knowledge, there has been no existing research study on exploring activities performed by the data collectors and analyzing their behavior dynamics that affect the quality of data collected for a mass-scale survey focusing on a religious mass gathering such as Hajj in Saudi Arabia, Kumbh Mela in India, etc. Additionally, no existing research study focused on scrutinizing the integrity of data collectors as well as proposed any technological intervention that can help in evaluating the integrity of data collectors.

### III. STUDY DESIGN AND DATA COLLECTION

To address vacancy in the literature, in this paper, we perform a rigorous study comprising a mass-scale survey. In this section, we describe the chronological development of our survey questionnaire, followed by our data collection process. Here, we adopt both online and in-person surveys to collect Hajj/Umrah experiences from the pilgrims. We disseminate the online version through emails and social media such as Facebook. On the other hand, we collect the in-person survey data through paid and unpaid data collectors who surveyed

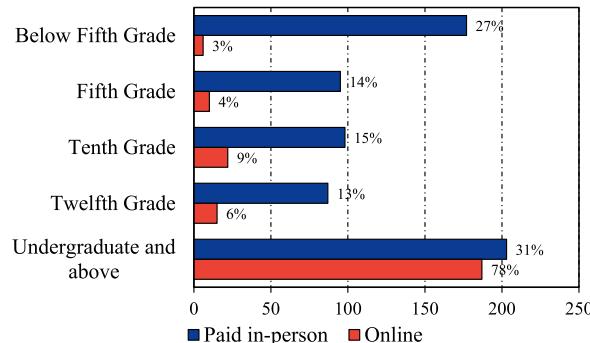
the Hajj pilgrims. Next, we describe the survey form preparation in detail.

#### A. SURVEY QUESTIONNAIRE PREPARATION

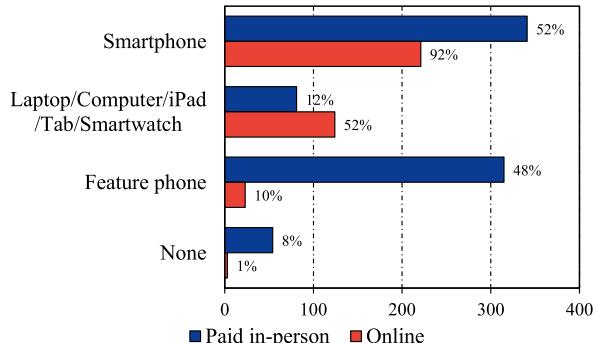
To figure out the problems that pilgrims face during Hajj/Umrah and their suggestions to resolve those problems, we prepare a questionnaire for pilgrims. The questionnaire was not anonymous. However, personal information (name, email, and contact number) was set as optional to respond. We collected the contact information as we had to contact some of the survey participants after the data collection process to verify the credibility of the collected data in some cases. We design this questionnaire as being worthy of a self-completed written form so that we can gather a large amount of data quickly from pilgrims simply by disseminating the forms and collecting responses from the pilgrims thereby. We organize our questions in a story-telling manner, which helps the pilgrims to remember their experiences of Hajj/Umrah.

We organize our questionnaire with 49 questions under 12 sections. At the beginning of the survey, we informed the participants that they might share their problematic or negative experiences (s) during their pilgrimage. Then, we collect demographic information (gender, age, literacy, country of living, language, etc.) from the pilgrims to correlate the pilgrims' difficulties with their demography. After getting the demographic data from a pilgrim, we collect information about difficulties faced pertinent to verbal communication during the pilgrimage. Next, we ask whether the pilgrim got lost during the pilgrimage or not. After that, we ask a pilgrim whether s/he faced any accommodation or transport problem during the pilgrimage. Then, we ask a pilgrim about the problems and possible solutions regarding Hajj/Umrah-related places such as Masjid al-Haram, Mina, and 'Arafah. We summarize the survey questionnaire in Table 1. We present details of the survey form including the survey form itself in English language in Appendix.

We design the Hajj survey questionnaire using the Google form service for online data collection purposes. Besides, we create a print-friendly version of our online survey questionnaire for in-person data collection purposes. Note that, we use an additional field in our in-person survey



(a) Highest educational qualification



(b) Electronic device usage

**FIGURE 2.** Demographic variations over survey participants observed in paid in-person survey and online survey.

questionnaire, i.e., form ID, to track the source of responses. To collect data from the local pilgrims, we translate our online survey questionnaire in Bengali through native Bengali speakers.

### B. RELIABILITY OF SURVEY QUESTIONNAIRE

We measure the internal consistency of the questionnaire to check its reliability. Among different statistical tests applicable for checking the internal consistency, we choose Kuder-Richardson 20 or KR-20 test [37]–[39]. As all the mandatory questions in the survey form are to be answered in a binary manner, we choose the KR-20 test in our case. For a sample size of 988 (which is the total number of participants in our survey), we get a KR-20 score of 0.62. This score is above 0.5, and thus, appears to be satisfactory from the perspective of reliability according to the study in [40]. This result confirms the reliability of the questionnaire used in our study. Further, please note that our study has been approved by the institution of the corresponding author.

### C. SURVEY FORM DISSEMINATION PROCESSES

After finalizing the survey questionnaire, we disseminate the survey questionnaire over the Internet to collect Hajj/Umrab experience from international pilgrims. Besides, we perform an in-person survey to collect such data from the pilgrims in Bangladesh and Saudi Arabia. Next, we describe both the modalities of our data collection. Note that we obtained institutional approval to conduct the study from the university of the first author.

#### 1) ONLINE DATA COLLECTION

The congregation in Hajj/Umrab consists of many people from different countries having diversified languages and cultures. To collect data on such types of congregation, we adopt a web-based survey to reach pilgrims from different countries within a short period. Accordingly, at first, we circulate our survey questionnaire in form of Google Survey through social media (Facebook) groups to different countries such as the USA, UK, Germany, Turkey, Australia, Estonia, etc., through

our friends and colleagues staying abroad. Most of those groups are private with few persons. Those groups are mostly student community groups in universities.

After that, we email our questionnaire to the faculty members of different universities in Afghanistan, Algeria, Australia, Bahrain, Egypt, India, Indonesia, Iran, Iraq, Jordan, Kenya, Kuwait, Malaysia, Nigeria, Pakistan, Saudi Arabia, South Africa, Sudan, Syria, Turkey, United Arab Emirates, and Uzbekistan. We collect their email addresses from their institutional web pages. Through this online data collection process advertised over both Facebook and email, we get 193 responses within six months from the pilgrims of 22 countries. To be more specific, we get 135 responses through advertising over Facebook and 58 responses through advertising over emails. Note that, we email the link of the survey questionnaire to 867 faculty members, however, we get only 58 responses resulting in a response rate of 6.69%. In addition to this English survey form, we circulate a Bengali translated Google Survey form over Facebook to collect data from the local pilgrims. We get 43 responses within one month in this way through the Bengali translated form.

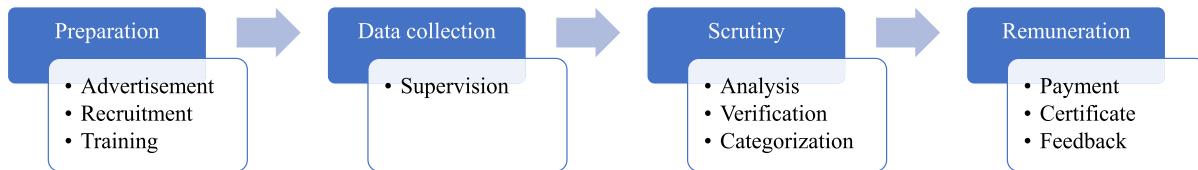
**Limitations:** The major drawbacks of online data collection are low response rates and demographic bias. Our online survey data shows that 83% of the participants have a university or higher education degree (Figure 2a). The reason behind this bias is that we circulate the online questionnaire among the students and teachers, as they are the most reachable group over the Internet. Besides, we have found that 92% of the participants of the online survey use smartphones for regular communication (Figure 2b). It implies that online surveys cannot reach the common people having little or no Internet access. To overcome the demographic bias problem and improve the response rate, we adopt an in-person data collection approach in addition to online data collection.

#### 2) IN-PERSON DATA COLLECTION

We collect the Hajj/Umrab experience from both Bangladesh and Saudi Arabia through in-person surveys. In Bangladesh,

**TABLE 2.** Summary of responses in our Hajj/Umrab survey data collection.

Survey mode	Voluntary or paid?	Language	Duration	# of responses	Responses per day	Coverage
Online	Voluntarily	English	6 months	193	1	22 countries
	Voluntarily	Bengali	1 month	43	1.4	4 countries
In-person	Voluntarily	English	1 month	40	1.3	2 countries
	Paid	Bengali	18 days	712	40	Bangladesh

**FIGURE 3.** Different stages of paid data collector management.

we first survey Hajj/Umrab pilgrims within the friends and family networks of the authors and get only 16 responses. Besides, we survey foreign pilgrims at Masjid al-Haram in Saudi Arabia during an Umrah pilgrimage by a few of the authors in February 2020. Since only a handful number of pilgrims can speak in English at Masjid al-Haram, we collect only 24 responses from there. The reason behind this small number of responses is the fact that, although in-person data collection is a credible source of data, it requires too much time and effort to survey a pilgrim.

Hence, we engage several data collectors to collect the experiences of Hajj/Umrab pilgrims from all over Bangladesh. For this purpose, we recruit data collectors. The recruited data collectors are mostly university students. The data collectors were staying in their homes situated in different parts of the country during the in-person data collection time of this study due to COVID-19 closure. The data collectors survey the pilgrims mostly in-person and in some cases over the phone. They mostly survey their relatives, neighbors, and acquaintance to ensure the credibility of survey responses. Using this approach, the response rate improves significantly and we get 712 responses within 18 days. We show the summary of responses collected through different modalities of our data collection in Table 2.

As the data collectors live in different parts of our country, we get many demographic variations in the survey responses. Our in-person data shows that only 32% of participants have a university or higher degree and 25% of participants have an educational qualification below 10<sup>th</sup> Grade (Figure 2a). In addition, we have found that 51% participants of the in-person surveys are smartphone users and 48% participants are feature-phone users (Figure 2a). Besides, we reach people having no Internet access, as 12.9% people of Bangladesh is yet to come under the coverage of Internet access [41].

For in-person data collection through paid data collectors, we follow some steps in paid data collector management. Here, we perform advertisement, training, data collection, and remuneration payment. We present details on each of these steps next.

#### D. PAID DATA COLLECTOR MANAGEMENT

To collect the experience of local Hajj/Umrab pilgrims, we recruit and train up some data collectors. Besides, we guide them in the whole data collection process. Later, we remunerate them based on their valid responses. We show the different stages of this process of collector management in Figure 3. We describe each of the stages in detail next.

##### 1) ADVERTISING IN SEARCH OF PAID DATA COLLECTORS

We prepare a registration form to recruit paid data collectors for our Hajj/Umrab survey. We mention that a data collector's task is to survey a Hajj/Umrab pilgrim face-to-face or over the telephone following our written questionnaire and then upload the response through a Google form. Based on the number of pilgrims surveyed by a data collector, we provide remuneration following a multi-level slabbed payment structure. Our payment rate for the data collectors is 60 BDT (0.71 USD) for each survey if the number of surveyed pilgrims is within the range of 1-20. The rate of payment gets higher if the number of surveyed pilgrims becomes more than 20. Accordingly, the payment rate becomes 65 BDT (0.77 USD) for 21-40 surveyed pilgrims and 70 BDT (0.83 USD) for more than 40 surveyed pilgrims. We describe this payment structure in the data collector registration form. Besides, we describe the approximate time required for each survey in the form. After that, we advertise the registration form on Facebook.

Within four days, 113 interested data collectors get registered in our portal. After that, we invite all data collectors through email to join a Facebook group for further communication. Among 113 registered data collectors, 98 data collectors join the Facebook group. To schedule the training session for data collectors, we create a Facebook poll with some probable time slots. Around 70 data collectors vote for their convenient time within two days. According to the highest voted option (34 votes), we schedule a training session over the Zoom platform [42]. After that, we notify the training schedule and the Zoom session link to the interested data collectors through emails and Facebook.

**TABLE 3.** Demographic traits of the data collectors.

Gender	Data collectors (n = 30)
Male	23
Female	7
<b>Age-range (years)</b>	
16-20	9
21-25	19
26-30	2
<b>Level of computer proficiency</b>	
Satisfactory	5
Good	17
Excellent	8

Religion	Data collectors (n = 30)
Islam	30
<b>Educational qualification</b>	
Undergraduate - 1st year	5
Undergraduate - 2nd year	6
Undergraduate - 3rd year	9
Undergraduate - 4th year	8
Graduate	2
<b>Prior experience in survey work</b>	
Yes	15
No	15

## 2) TRAINING EVENT FOR DATA COLLECTORS

We conduct a two hours long interactive training session on our Hajj/Umrab survey. Among 113 registered data collectors, 65 data collectors (58%) participate in the training session. In the training session, we explain the questionnaire in detail to the data collectors, demonstrate a complete submission process of the survey, and respond to all questions from the participants of the training session. Moreover, we provide some helpful techniques to collect survey data from pilgrims based on our previous experience. Finally, we describe the terms and conditions of our survey data collection. As Hajj/Umrab is of utmost spiritual significance, we adopt a zero-tolerance policy to any breach in our code of conduct of the survey to ensure the highest possible integrity over our collected survey data. If we detect a single invalid form from a data collector, we discard all forms collected by that data collector. We present all these adopted stringent policies to the participants of the training session so that they can get aware of it. After the training session is over, we email the Hajj/Umrab survey form link, the recorded video of the training session, written terms and conditions mentioning all our policies, and required additional materials to all the interested data collectors.

## 3) DATA COLLECTION

We collect responses from the data collectors, i.e., their collected surveyed information covering experiences of pilgrims during Hajj/Umrab, through a specified Google form. Each data collector has to provide a unique ID for each response for tracking the data collector later. Besides, the email address of the submitting data collector is mandatory during the submission process to facilitate the tracking. After submitting a response, an editable copy of the response is sent to the data collector automatically. Initially, we accept 535 responses within ten days. After that, we pause the data collection for two days and perform some initial checks on the response frequency for each data collector, variation over data in the submitted responses for each data collector, etc. After two days of break, we resume the data collection by announcing a gentle warning in our Facebook group about potential fake data in the submitted responses. At this stage, we receive 178 data within the next six days. In total, we receive 712 responses from 53 data collectors over a total period of 18 days.

## 4) REMUNERATION AND FEEDBACK

Using different quantitative and qualitative methods (described later), we identify three deceptive data collectors, who severely violate the survey protocol. Excluding them through having explicit notification with detailed justifications to them, we remunerate all the other data collectors and also grant them digital certificates recognizing their effort. After that, we conduct a feedback survey to know their experiences about this data collection process such as problems faced during data collection, problems and suggestions regarding the survey questionnaire, and their comments on training and management. 30 data collectors filled up the feedback survey. In the next subsection, we will discuss the demography of those data collectors who filled up the feedback survey.

## 5) DEMOGRAPHY OF DATA COLLECTORS

A total of 53 data collectors participated in the data collection. Among them, we could collect the demography of 30 data collectors through the feedback survey, as the other 23 data collectors did not participate in the feedback survey. We show the demography of those 30 data collectors in Table 3. Notably, most of the data collectors are undergraduate students. This is because the procedure of gathering the data collectors was through different university students, groups covering diverse locations all over the country. Most of them have a good level of computer proficiency and many also had prior experiences in survey work. Another important insight is that all the data collectors are Muslim. This is due to the nature of this work, where the data collectors had to talk to Muslim pilgrims, and most importantly a baseline prior knowledge about Hajj rituals is expected for the work. Though a person in the religion of Hinduism contacted to participate as data collectors, after knowing details about the mode of our survey, he voluntarily drew it out on his own.

## E. RESEARCH POSITIONALITY STATEMENT

In this paper, all of the authors are Muslim by birth. Besides, all of them are Bangladeshi except one, who is a Saudi Arabian. Two of the authors have direct experiences on performing Hajj pilgrimage, and five of the authors have direct experiences on performing Umrah pilgrimage. During a very recent Umrah pilgrimage in February 2020, three of

**TABLE 4.** Conflicting cases and their severity in collecting survey responses.

Conflicts	Conflicting cases (Question ID: Response)	Conflicting conditions	Remarks	Severity
Conflict 1	Q38: Perform only Umrah Q48: Got heatstroke, sickness at Arafah	Q38 and Q48	Might not applicable for Saudi Arabian citizens	Very high
Conflict 2	Q38: Perform only Umrah Q48: Transportation problem from 8-13 Dhul-Hijjah	Q38 and Q48	Might not applicable for Saudi Arabian citizens	Very high
Conflict 3	Q13: Faced no difficulty in communication Q39: It is challenging to communicate with Mina volunteers	Q13 and Q39	Depends on the sequence of the questions	High
Conflict 4	Q13: Faced no difficulty in communication Q19: Language barrier as a reason for getting lost	Q13 and Q19	Depends on the sequence of the questions	High
Conflict 5	Q21: Did not get sick after/when got lost Q22: Did not get sick during Hajj/Umrah Q48: Got sick in Arafah	Q21 and Q22 and Q48	Depends on the sequence of the questions	High
Conflict 6	Q16: Did not get lost in Arafah Q21: Got sick after/when got lost Q22: Did not get sick during Hajj/Umrah Q48: Got sick in Arafah	Q16 and Q21 and Q22 and Q48	Depends on the sequence of the questions	High
Conflict 7	Q21: Did not get sick after/when got lost Q22: Did not get sick during Hajj/Umrah Q50: Got sick several times (more than Zero)	Q21 and Q22 and Q50	Depends on the sequence of the questions	High
Conflict 8	Q21: Got sick after/when got lost or Q22: Got sick during Hajj/Umrah Q50: Did not get sick at all (Zero)	(Q21 or Q22) and Q50	Depends on the sequence of the questions	High
Conflict 9	Q11: Use only a feature phone Q48: Can not use WiFi at the hotel	Q11 and Q48	Collective experience may differ	Medium
Conflict 10	Q8: Can speak Arabian or Hindi or Urdu Q39: Can not take help from surrounding people due to the language barrier	Q8 and Q39	1. Hindi and Urdu may not be applicable everywhere; 2. Arabic might have been learned after performing Hajj/Umrah	Medium
Conflict 11	Q8: Can speak Arabian or Hindi or Urdu Q13: Faced difficulty in communication	Q8 and Q13	1. Hindi and Urdu may not be applicable everywhere; 2. Arabic might have been learned after performing Hajj/Umrah	Medium
Conflict 12	Q12: Hajj in winter (October - January) Q48: Got heatstroke, sickness at Arafah	Q12 and Q48	Depends on the age of the pilgrim	Medium

the authors had surveyed pilgrims in person at the Masjid al-Haram in Mecca, Saudi Arabia. Besides, all of the authors except one had surveyed the pilgrims in person among their relatives, neighbors, and acquaintances in Bangladesh. Nonetheless, the research study has been approved to be pursued by the Bangladesh University of Engineering and Technology.

#### IV. ANALYSIS ON DATA COLLECTION PROCESS

After finishing the in-person data collection process, we carry out aggregated analysis on the data collected by data collectors, we perform collaborative scrutiny through focused group discussions by expert reviewers to evaluate the integrity of the data collectors. Here, the notion of ‘expert reviewers’ represents project scientists or experienced data collectors as adopted in an existing research study [1]. In our study, we (the authors only) scrutinize the data collection process. For this purpose, we contact both pilgrims and data collectors to resolve the potential anomalies found in the data. Note that, we analyze the in-person survey data only in this way. Next, we present the analysis of over-collected in-person data in detail.

##### A. AGGREGATED DATA ANALYSIS

We perform different aggregated data analyses over the responses collected by each data collector such as

percentages of blank fields, repetition of data, and conflicting answers. Besides, we also perform statistical analysis and learning-based analysis. In the data under analyses, blank fields represent no difficulty, illness, or obstacles faced by the pilgrims. Repetition means noticeable similarity among all responses collected by a data collector from different pilgrims. Besides, conflicting cases refer to inconsistency in a collected response. Now, we elaborate more on our conflict analysis as well as statistical analysis and learning-based analysis.

##### 1) CONFLICT ANALYSIS

Conflict analysis represents discrepancies within the survey response of a particular pilgrim. For example, if a pilgrim does not mention being sick at the initial phase of the questionnaire (Q22), however, s/he mentions illness in the latter part (Q48 and Q51), this appears to be a potentially conflicting response. We identify 12 such potential different conflicting cases among the responses collected in the survey. We classify them into three classes based on the possible severity of being conflicting, i.e., very high, high, and medium. Table 4 shows details of all the conflicting cases. Here, Conflict 1 and 2 are considered as very-highly severe potential conflicts. These occur whenever a pilgrim says that he performs only Umrah but gets sick in Arafah

or faces transportation problems during 8-13 Dhul-Hijjah, which should be impossible as rituals in Umrah do not cover the place Arafah and the time 8-13 Dhul-Hijjah. On the other hand, Conflict 3, 4, 5, 6, 7, and 8 are considered highly severe potential conflicts. These may occur due to the sequence of the survey questionnaire, i.e., pilgrims often cannot remember the minute details of their Hajj/Umrah experiences at the beginning of the survey and can do it later as the survey progresses. Conflict 6 and 7 are considered medium severe potential conflicts. These conflicts occur when a pilgrim can speak in Arabic/Urdu/Hindi language, however, s/he faces communication problems with the surrounding people. This should not happen as people knowing Arabic/Urdu/Hindi do not generally face communication problems. We consider its severity as a medium, since the pilgrim may have learned the language after performing Hajj/Umrah.

## 2) STATISTICAL AND LEARNING-BASED ANALYSES

Since most of the survey questions have categorical responses, we adopt an entropy-based approach to measure the level of information present in the responses [43]. To formulate entropy, let, the number of possible events is  $n$ , and the discrete probability of event  $i$  is  $p_i$ . Thus, we can calculate the information for a random variable X. This might be written as a function  $H(X)$  which is elaborated in the following equation.

$$H(X) = H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

In entropy, the level of information present in X increases as  $H(X)$  increases. We first consider each response to calculate the entropy. In each response, there are  $n = 50$  questions to be answered. We calculate the entropy on the overall responses covering all these questions. Then, we average the values for each of the data collectors in addition. To elaborate more with an example, suppose, a data collector collected 20 responses. We calculate entropy for each of the responses and then take an average of them. Thus, we calculate the average entropy for each of the data collectors.

In addition to entropy-based analysis, we adopt an unsupervised learning approach to measure the information content of a response. We measure the information content based on the number of blank responses, the number of “Do not know” responses (DK responses [44]), and entropy. Here, “blank responses” refer to responses with empty or no string. This might happen when a participant said that s/he did not face a problem while responding to a question on whether s/he faced any problem in her/his pilgrimage. Besides, “Do not know” responses refer to responses where a participant said that s/he did not know anything about an incident that s/he is being asked for to share her/his experience during the pilgrimage. Nonetheless, as already mentioned, entropy represents the variation in the responses. We consider these three measures as indicators of the level of information in a response. Based on the level of information we find two classes of

information, i.e., “More informative” and “Less informative”. If a response has overall fewer blank responses, fewer “Do not know” responses, and less entropy, we consider the response as a potential “More informative” response and vice versa. Finally, we apply the k-means clustering algorithm [45] and classify the responses into “More informative” and “Less informative” clusters accordingly. We will describe the formation of clusters in detail in Section V-D.

## 3) RESPONSE VERIFICATION

After analyzing the responses from various aspects, we sort out potentially conflicting responses having high percentages of blank fields, repetitions, and conflicts. We identify these pilgrims and have several recorded phone conversations with them using their phone numbers provided in the survey responses. We first ask the pilgrims whether anyone surveyed them on their Hajj/Umrah experiences or not and the average time duration of the survey. After that, we ask them about the identified inconsistent answers in their responses and check whether the pilgrims’ conversations with us match with the survey responses. If we find that the conflicting answers or blank fields are not from the pilgrims, rather they have been inserted by the data collectors, we add the corresponding data collectors to our anomalous-list or black short-list. Next, we discuss how we find anomalous cases through our focused group discussion.

## B. FOCUSED GROUP DISCUSSION

In our focused group discussion, we contact the short-listed data collectors and have recorded phone conversations to clarify the inconsistent issues. We notified the data collectors that the conversations are being recorded. We ask the data collectors about whether s/he collected first-hand data or not and the average duration of all their surveys. Besides, we request clarifications on dissimilarities between the information provided by pilgrims and information submitted by the data collectors, degree of alignment with the survey protocol, and opinions regarding the high percentages of blank fields or repetitions or conflicts in their submitted responses. If we find data manipulation, misinterpreting survey questionnaires, or survey protocol violation, we label the corresponding data collector as an “unreliable” one. Note that, in case of additional discrepancies, we contact both the pilgrims and the corresponding data collectors for further clarifications. Only after having convincing clarifications, we classify a data collector as either “reliable” or “unreliable”. After such thorough investigation, among the 53 data collectors, we find 29 data collectors as “reliable”. Accordingly, we finalize the ground truth of the dataset. Here, each instance of the ground truth dataset has two entries - data collector id and label (either “reliable” or “unreliable”).

## V. FINDINGS

In this section, we describe the results of our aggregated data analysis and focused group discussion. Besides, we answer

the aforementioned research questions according to the obtained results.

#### A. RQ1: ACTIVITIES OF HUMAN DATA COLLECTORS

As per our phone conversations with both pilgrims and data collectors, we find diversified activities performed by the data collectors during the process of data collection. These activities can be broadly classified into two types of data collections namely first-hand data collection and second-hand data collection. First-hand data collection includes a one-to-one in-person or telephone (phone/social media) survey. During a telephone survey, some of the data collectors provide the soft copy or screen-shots of the survey form, however, the form is filled up by the data collectors. On the other hand, second-hand data collection includes getting responses using someone else, providing a hard copy of the form to pilgrims to get it filled up by the pilgrims, or advertising the form in social media channels and let the pilgrims fill-up the form in the absence of the data collector. Most of the second-hand data collectors get the spouses' (mostly wife) responses from the counterpart (mostly husband) through verbal communication with the counterpart. The reason behind such activities is the social context prevailing in Bangladesh. Here, it is often common to take a wife's response through her husband, since it is not allowable for the wife to contact others for religious obligations. Therefore, socio-geographic factors, such as gender barriers during communication in rural areas, hinder data collectors to collect first-hand data from rural female communities.

In addition to second-hand data, some of the data collectors do not communicate with the pilgrims at all for collecting responses and provide fake data. Some data collectors ( $n = 3$ ) deceitfully give similar responses without talking to pilgrims. One of the data collectors opens another survey website where pilgrims are invited to fill up the forms in exchange for a lower payment compared to what we advertised. The data collector wants to collect a higher amount of responses for monetary gain in this manner. We confront this deceptive data collector through the information provided in his/her own survey website. We further confirm these over a recorded phone conversion with him/her. To exemplify, the deceptive activity can be understood from a part of the conversation as follows:

*"Let me explain to you. I have an aunt in the neighborhood. I asked her son, whether I can collect data from her (for the survey). Her son said, "Give us the survey form, we will take care of that". I said, "I am unable to do so." Then he asked for the questionnaires. Therefore, I build a website with the questionnaires to collect data from his mother."*

Despite sincerity towards the designated task, some data collectors provide flawed data due to misunderstanding the survey questions. One of the data collectors expresses his/her honest confession regarding submitting flawed data in this way as follows:

*"The form IDs that I think maybe problematic are 31, 32, 33, 35, 36-39, and 42-48-basically from above 30 as far as*

*I can remember because of the way I took the responses. I humbly apologize for breaking the terms and conditions of the survey, for not being able to collect data properly, and for not being able to take only from those who are truly willing to give the proper information. If my data is not helping in this survey, you can eliminate my data. Thank you"*

Besides, since our survey focuses on a religious event, some of the data collectors consider this data collection process as a religious duty and do not want to make their payments. One of them expresses his/her feelings about this religious activity as follows:

*"I don't want the payment. In Shaa Allah, I'll take the official participation certificate in the Hajj Survey. I'm grateful to the Almighty to be a part of such an activity. I have a cordial request to you and the team behind this survey-if further such activities are conducted, whether voluntary work or academic work, please let me know if you think I'll be of any help."*

Thus, we understand that religious aspects influence positively to escalate integrity of some of the data collectors. Additionally, participants are observed to share their religious experiences with higher interest, which eventually helps data collectors in getting more informative data. In addition, the influence of other socio-geographic aspects, such as, the reluctance of urban participants to contribute data in the survey, is also noted from the feedback of our data collectors. Here, as per our data collection experiences, the urban community maintains a higher amount of professional busy times compared to that of rural people. Hence, the urban community is less likely to show enthusiasm in contributing data to the survey process. We find all these aspects of our survey during our focused group discussions. During the discussions, we experience several influential factors that we elaborate on next.

#### B. RQ2: INFLUENTIAL FACTORS FOR COLLABORATIVE SCRUTINY

While scrutinizing the activities of data collectors during the data collection process, we consider several factors to classify a data collector as reliable or unreliable. Some influential factors are as follows: the degree of alignment with our survey protocol, outcomes of aggregated analysis, conversation with both participants and corresponding data collectors, data collectors' response rates, the extent of data collectors' efforts, and participants' religious contexts. We consider these factors to categorize the data collectors during our focused group discussions. Next, we summarize the categorization process along with justifications.

##### 1) UNRELIABLE DATA COLLECTORS

If any data collector provides fake data or publishes our form on the website or social media, we mark him/her as an unreliable data collector. As the penalty of their forgery, we discard all of their responses and make no payment to them. We identify three such deceptive data collectors. We also mark a second-hand data collector as an unreliable one, since it directly violates our survey protocol. In our written terms

**TABLE 5.** Categorization of the data collectors and their payments.

Category	Payment	Major cause	# of data collectors	% of data collectors
Unreliable	Full	Low anomalous data	30	57.6
	Full	Second-hand data collection	17	32.1
	Partial	High anomalous data	3	5.7
	No	Deception	3	5.7
Total			53	100

**TABLE 6.** Overall conflicts in all modes of data collections.

Quantity	Data collected through				Total	
	Online survey		In-person survey			
	Locally	Internationally	By the authors	By the paid data collectors		
Total # of responses	43	193	40	712	988	
# of responses with conflicting cases	10	33	5	124	172	
% of responses with conflicting cases	23.3	17.1	12.5	17.4	17.4	
	31.6		17.2			

and conditions, we explicitly mention that the acceptable data collection methods are either in-person data collection using a hard copy of the survey form or surveying pilgrims over the phone/social media conversation. We enforce these to make sure one-to-one correspondences between the data collector and participant leaving no or the least chance of having any communication gap. Even after mentioning these, we identify 17 second-hand data collectors. In this case, we do not discard all of their responses. Rather, we use their responses excluding the second-hand data in our analysis. We also make full payment to them to remain ethically sound to our data collectors. This ethical issue is also raised by a second-hand data collector:

*“Since some of my data recognized as invalid and as a result, all of my data will be cancel out, you should ensure me you don’t use anyone data for your research purposes which I submitted to you.”*

Thus, as we make the full payments to the second-hand data collectors, this ethical issue gets well addressed. Accordingly, we can use the data collected by such data collectors except for the second-hand ones. Apart from these types of data collectors, we penalize three data collectors and make partial payments to them due to higher percentages of blank fields and conflicting cases in their submitted responses, and misinterpreting the questionnaire. Finally, in total, we filter out 23 unreliable data collectors. We show the categorization of these data collectors as well as other ones in Table 5.

## 2) RELIABLE DATA COLLECTORS

We categorize a data collector as a reliable one if we get a small discrepancy in terms of repetition of data, conflicting cases, and statistical checks. Note that, a high number of blank fields or repetitions does not always indicate suspicious data. If a pilgrim performed Umrah only, s/he might face difficulties in Masjid al-Haram only resulting in a higher number of blanks in many other parts in our survey. Besides, recording responses from a group of pilgrims who performed Hajj/Umrah together could result in substantial repetitions

of data. Additionally, we do not categorize a data collector as unreliable and do not penalize the data collector if inconsistent data get submitted due to the participant pilgrim’s misunderstanding as to his/her inconsistent response thereof. Following all these approaches, at the end of our collaborative scrutinizing, we get 30 reliable data collectors in total.

## C. RQ3: EFFICACY OF CONFLICT ANALYSIS

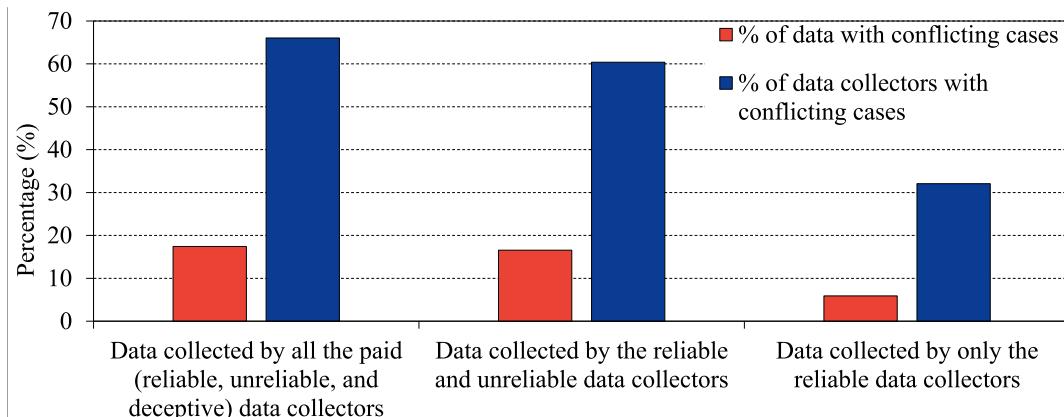
To measure the efficacy of our information-based conflict analysis, we analyze both online survey data and in-person survey data from different perspectives. We discuss the outcomes of the analysis in detail next.

### 1) ANALYSIS OVER ALL SURVEY DATA

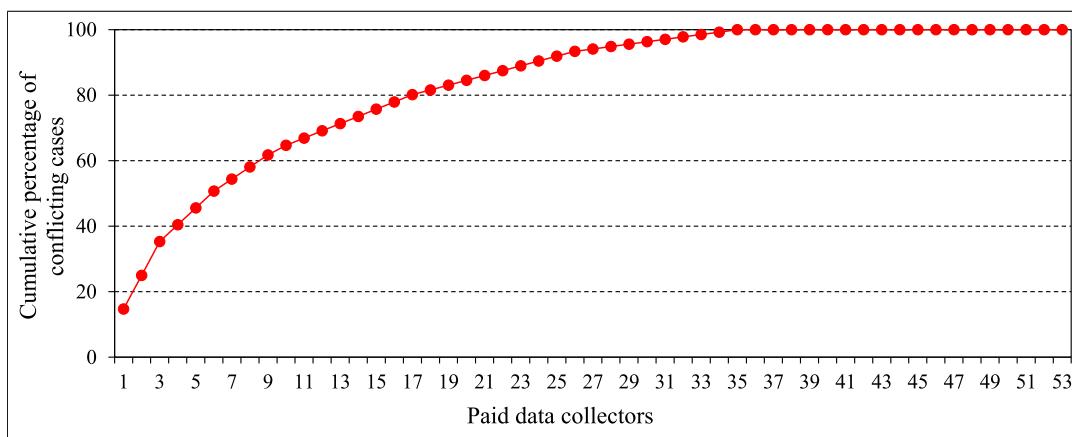
We present the overall conflicts in all modes of survey data collection in Table 6. We can see that the data collected by the authors through in-person survey have the lowest rate (12.5%) of conflicting cases and the data collected locally through online survey has the highest rate (23.3%) of conflicting cases. The data collected internationally through the online survey and the data collected by the data collectors through in-person survey have a similar rate of conflicting cases (17.1% and 17.4% respectively). Besides, from an aggregated perspective, the rate of conflicting cases over the data collected through online survey (31.6%) is more than that found over the data collected through in-person survey (17.2%). These values suggest that, utilizing online survey methodology leads to more conflicting cases in the collected data, which in turn implies having less reliable data through online survey. Thus, in a large-scale survey such as Hajj pilgrimage, the in-person survey should appear to be more effective in collecting more reliable data than the online survey.

### 2) ANALYSIS OVER MANUALLY-COLLECTED SURVEY DATA

During collaborative scrutiny, we find some data collectors who provide fake data intentionally in our survey submission system. We consider these data collectors as deceptive ones.



**FIGURE 4.** Percentages of data and data collectors with conflicting cases.



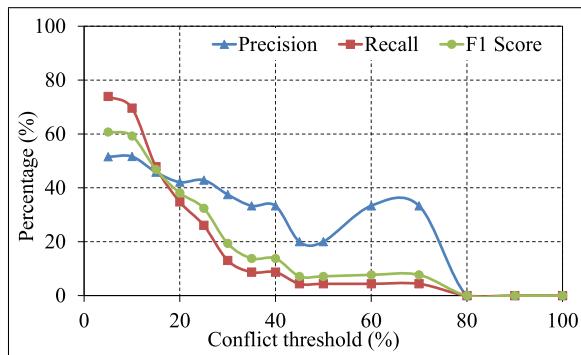
**FIGURE 5.** Cumulative percentages of conflicting cases with all paid data collectors.

Besides, some data collectors appear to be unreliable from several different perspectives as stated previously. Figure 4 presents the percentages of data and data collectors having conflicting cases in different subsets of our dataset, i.e., overall data collectors, over reliable and unreliable (excluding the deceptive ones) data collectors, and only the reliable (excluding unreliable and deceptive ones) data collectors. We can see that the data collected by the data collectors contain 17.4% conflicting cases, and these conflicting cases were collected by 35 (66%) data collectors. After excluding the fake data provided by the deceptive data collectors, we have 16.6% data having conflicting cases, and these are collected by 32 (60%) data collectors. This figure also shows that after eliminating all the data from the deceptive and unreliable data collectors, the dataset contains only 6% conflicting cases and these are collected by 17 (32%) data collectors.

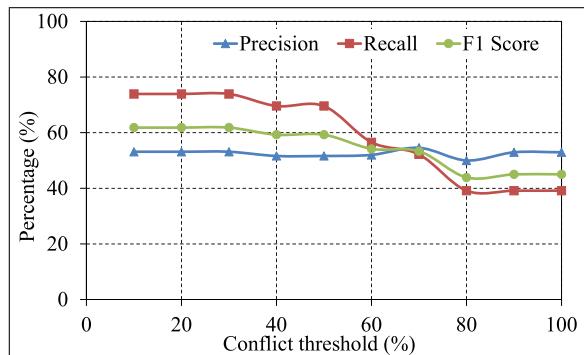
Figure 5 presents the cumulative percentage of the conflicting cases with all paid data collectors (53). It implies that only 4% of the data collectors contain 25% of the conflicting cases, around 11% of the data collectors contain around 51% of the conflicting cases, 28% of the data collectors contain around 76% of the conflicting cases, and

around 66% of the data collectors contain 100% of the conflicting cases.

To analyze the effectiveness of conflict analysis, we consider two metrics separately - the percentage of conflicting cases in a data collector's data, and the percentage of the highly severe conflicting cases (Section IV-A1) over all the conflicting cases in a data collector's data. Besides, based on the outcomes of conflict analysis, we label each data collector according to the percentage of conflicting cases in the survey data s/he collected. Here, we label a data collector as “reliable”, if the percentage of conflicting cases over all the data collected by the data collector is within a particular threshold. Conversely, we label a data collector as “unreliable”, if the percentage of conflicting cases over all the data collected by the data collector exceeds the threshold. In this regard, we take into account different thresholds (over the percentage of conflict) such as 5%, 10%, 15%, ..., 95%, and 100%. Accordingly, for each of these thresholds, we compare the label of each data collector assigned in this way against that obtained from corresponding ground truth label. Based on this comparison, we further calculate statistical accuracy measures such as precision, recall, and F1 score.

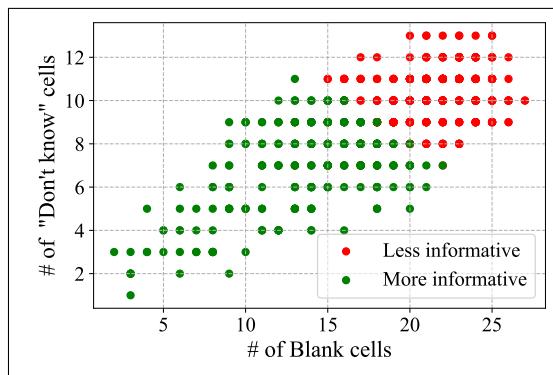


(a) Variation in accuracy in response to change in threshold for percentage of all "conflicting cases"

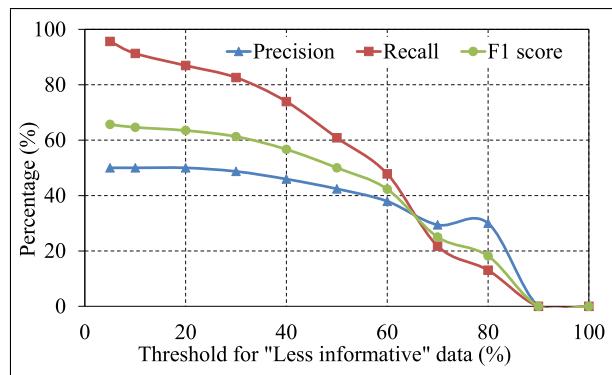


(b) Variation in accuracy in response to change in threshold for percentage of only "highly severe conflicting cases"

**FIGURE 6. Accuracy of conflict analysis in identifying unreliable data collectors based on two different metrics.**



(a) Clusters of responses based on information content in each response



(b) Variation in accuracy in response to change in threshold for percentage of "Less informative" data

**FIGURE 7. Formation of clusters and accuracy of clustering-based analysis in identifying unreliable data collectors.**

These accuracy measures imply how good the conflict analysis approach is to detect unreliable data collectors.

We present outcomes of calculating the accuracy measures in Figure 6. Here, Figure 6a presents the variation in the accuracy measures in response to changing the threshold. Besides, Figure 6b also presents the variation in the accuracy measures in response to changing the threshold, however, through being limited only to the "highly severe conflicting cases" that has already been defined in Section IV-A1. Now, recall values in Figure 6a and 6b depict that, at the value of 5% for the conflict threshold, 74% of the unreliable data collectors can be identified through using any of the metrics. Additionally, the recall values appear to be better in the case of considering threshold only on "highly severe conflicting cases". Similarly, the precision values also appear to be better in the case of considering threshold only on "highly severe conflicting cases". As a result, we can achieve better F1 scores while considering threshold only on "highly severe conflicting cases". Thus, our conflict analysis considering threshold only on "highly severe conflicting cases" offers a good indication of unreliable data collectors.

#### D. RQ4: EFFICACY OF LEARNING-BASED ANALYSIS

According to statistical analysis, we get a very small range of entropy on overall responses, (minimum = 3 and maximum = 4.9), which is not sufficient to classify more and less informative responses. On the other hand, in clustering-based analysis, we get two clusters of responses, i.e., "More informative" (# of responses = 396) and "Less informative" (# of responses = 316). In Figure 7a, we show the scatter plot of both the clusters where green and red points represent "More informative" and "Less informative" responses respectively. Note that, as there exist many overlapping points, it appears to have fewer data points in the figure compared to the 712 responses in total.

After the formation of the clusters, we evaluate each data collector according to the percentage of "Less informative" responses s/he collected. We do so as we find a high percentage of blank or "Don't know" responses for "Less informative" ones in case of unreliable cases as already mentioned. We reveal this aspect from our focused group discussions. Accordingly, we label a data collector as "reliable", if the percentage of "Less informative" data over all the data

collected by the data collector is within a particular threshold. Conversely, we label a data collector as “unreliable”, if the percentage of “Less informative” data spans over all the data collected by the data collector that eventually exceeds the threshold. In this regard, we take into account different thresholds (over the percentage of “Less informative” data) such as 5%, 10%, 15%, ..., 95%, and 100%. Accordingly, for each of these thresholds, we compare the label of each data collector assigned in this way against that obtained from the corresponding ground truth label. Based on this comparison, we further calculate statistical accuracy measures such as precision, recall, and F1 score. These accuracy measures imply how good the learning-based analysis approach is to detect unreliable data collectors.

We present outcomes of calculating the accuracy measures in Figure 7b. Here, Figure 7b presents the variation in the accuracy measures in response to changing the threshold. The figure also depicts that, at the value of 5% for the “Less informative” threshold, the recall value is 99%. Thus, 99% of the unreliable data collectors can be identified through clustering-based analysis which is a huge improvement compared to earlier conflict analysis (recall is 74%). However, in this case, precision is 47%, which implies that this approach can detect “reliable” data collectors as “unreliable” ones engendering some false positive cases. Despite this, the F1 score of 64% is still an improvement over the conflict analysis that has an F1 score of 62%.

## VI. DISCUSSION

Our field study and subsequent analyses reveal that *data collectors exhibit a noteworthy probability of performing their tasks deceitfully and unreliably*. We observe that around 43% data collectors adopt unreliable approaches to perform their designated tasks, among which confirmed misconducts having ill-motive along with treacherous acts are observed in case of 6% of the data collectors. The tasks having ill-motive, deceitful, and/or unreliable natures cover different spectrums. Examples include creating a new data collector group with lower payments violating data collection guidelines, providing highly repetitive false responses to increase own payment claim, taking forbidden advantages of onboard-ing online communities to collect secondary data and then place them as one-to-one collected data, etc.

### A. IMPACTS OF RELIGIOUS AND SOCIO-GEOGRAPHICAL ASPECTS

Our study reveals that *religious and socio-geographical aspects can exhibit both favorable and unfavorable impacts on the data collection process of a mass-scale survey*. Here, we observe that data collectors show high interest in taking part in a religion-related survey. The flavor of religion in the survey process is often observed to encourage increased integrity over the data collectors. Besides, we find participation from the survey participants to be more cordial in the process of sharing own *religious* experiences, which results in more informative data in the survey. Therefore,

we learn through our study that having a religious flavor in the survey process tends to assist the data collection process. On the other hand, different socio-geographical aspects such as gender gap in communication with survey participants in rural areas, professional busy times of the survey participants in the urban areas, etc., are observed to create barriers during the data collection process. In the case of the gender gap, the female community is discouraged to share their data through direct communication with male data collectors in rural areas owing to religious as well as social practices. Besides, in the case of professional busy times, job holders are often reluctant to participate in interviews due to having professional busy times and this happens mostly in the urban areas. Apart from these aspects, incentive-based data collection appears to be another cause of adopting deceitful and unreliable approaches by the data collectors.

### B. IMPORTANT FACTORS TO BE CONSIDERED DURING COLLABORATIVE SCRUTINY

We find that *there exist some crucial factors that need to be considered with due diligence during collaborative scrutiny of the integrity of data collectors*. Some of the notable factors in this regard include analyzing the degree of alignment of the data collectors with the set survey protocol, individual response rates of data collectors, pattern of one-to-one conversation between survey participants and the data collectors, correctness of understanding or perception of the data collector regarding each of the survey topics, etc. Systematic observation as well as in-depth analysis over these factors within focused group discussion of expert reviewers guides to identify possible cases of loss of integrity and unreliable data collection. Such focused group discussion and its successful outcomes hint towards the necessity of human-in-the-loop. This notion of human-in-the-loop becomes unavoidable in case of an absence of a full-fledged automated analysis to confirm the identification of loss of integrity and unreliable data collection. Thus, it becomes necessary to identify whether the full fledges automated analysis is available or not.

### C. SUFFICIENCY OF AUTOMATED ANALYSIS IN MEASURING INTEGRITY OF DATA COLLECTORS

Through our rigorous analysis, we demonstrate that *automated analysis alone is not sufficient to measure and confirm the integrity of data collectors*. For example, in our learning-based analysis, we had an entropy metric that considers repetitive responses as an indicator for identifying possible anomalous data collectors. However, the repetitive response may also get generated naturally, as a group of pilgrims who performed their pilgrimage in a single group may face the same kind of problems. Besides, some data collectors collect responses from family members who performed Hajj/Umrah together and participated in our survey at the same time and location. In this situation, participants mostly tend to mention the same kind of problems and tend to provide similar suggestions for solving the problems. This happens as the responses can get influenced by listening from the other

members of the family who also participate in the survey. *This is why manual intervention appears to be unavoidable following an automated analysis to precisely identify possible cases of unreliable data collection.* Nonetheless, our proposed techniques of conflict analysis and learning-based analysis can present a shortlist of possible unreliable data collectors based on their collected data. Though it can effectively identify all the unreliable data collectors, it also includes some reliable data collectors in the shortlist. As such, to filter out only the unreliable data collectors from the shortlist, subsequent collaborative manual scrutiny becomes compulsory portraying the potentially unavoidable necessity of human-in-the-loop.

#### D. IMPLICATIONS OF OUR STUDY

Our study aims to understand the scrutinizing factors that can help evaluate the integrity of data collectors in the case of a mass-scale survey on the experience of a religious event having a mass gathering. Our findings of this study will help administrations, survey organizers, and researchers to get a comprehensive idea about how to evaluate the integrity of data collectors so that reliable and unreliable data collectors can be distinguished. In addition, this study proposes a computational method to identify potentially unreliable data collectors based on their collected data, which can assist in an initial shortlisting of potentially unreliable data collectors. This will reduce the demands on effort and cognitive load to manage data collectors as well.

Further, we observe that religious and socio-geographical aspects play a major part in the process of executing tasks by the data collectors. For example, people often enthusiastically share their religious experiences resulting in more informative survey records. Thus, religious factors can assist data collectors to collect more informative data. Besides, some of the data collectors are observed to deny taking any incentive for their tasks of data collection, since they think that taking part in a religion-related work should not have any worldly or monetary gain. Thus, religious aspects exhibit motivation among some of the data collectors as well in executing their tasks with increased integrity and effort. Nonetheless, some socio-geographical factors such as the gender gap in rural areas, professional busy times in urban areas, etc., hinder the process of data collection. These hindrances in turn limit collecting planned responses such as responses from female participants in the rural area, responses from professional people in the urban areas, etc. These findings should facilitate planning a survey on road to ensuring sufficient timely data collection confirming diversity over the participants.

Our findings suggest that accurate and conclusive scrutiny over the integrity of data collectors necessitates the collaborative participation of expert reviewers. Here, the review process needs to adopt a standard methodological approach, where cross-validation of recorded data by directly communicating with some of the survey participants, double-checking the alignment of the data collectors with the set survey protocol during and after the data collection process, supervising the data collection process and solve any issue that comes

in real life, and other relevant tasks need to be performed. Besides, some important aspects such as consistency of information, similarity in responses collected from different participants, etc., need to be thoroughly reviewed during the scrutiny process. As such, survey organizers need to adopt these approaches during manual scrutiny of the data collectors.

Additionally, conflict analysis and learning-based analysis exhibit a noteworthy efficacy in the process of scrutinizing data collectors. However, our findings suggest that these computational methods alone are not sufficient for performing the scrutinizing task precisely. This happens as, although the computational methods prepare a shortlist of potential deceivers, manual scrutiny with collaborative interactions is needed to explicitly identify actual unreliable data collectors from the shortlist. This finding will help survey organizers design an efficient mixed method of semi-automated (combining both manual and automated processes) scrutiny process through which unreliable data collectors can be filtered out.

Finally, this study explores the ways of analytical scrutiny of integrity of data collectors, who collect data from participants for a mass-scale survey. It provides a foundation for future work on evaluating the efficacy and applicability of our proposed scrutiny process for other types of mass-scale surveys covering the diversified spectrum of use cases such as other religious surveys, national-level surveys, business surveys, etc.

#### VII. CONCLUSION

To figure out the nature and severity of diversified difficulties that pilgrims experience during massive religious congregations (such as Hajj and Umrah), surveying the pilgrims is unavoidable. In this study, we perform such a mass-scale religious survey for Hajj/Umrah using primarily two different modalities - in-person and online. Here, we perform the former mode of data collection through employing two different types of data collectors - voluntary and paid. We recruit, train, manage, and scrutinize the paid data collectors following a particular methodical process. We make them prepared for performing the survey task through the process. We critically analyze the paid data collectors' overall data collection procedures and activities as well as figure out different opportunities and challenges while performing the whole survey process. At the same time, we experience and learn different attributes of reactions from the data collectors throughout the survey process. We scrutinize the integrity of our data collectors based on their collected data through collaborative scrutiny and focused group discussion by expert reviewers. We find out different factors that influence the collaborative scrutiny of our data collectors. We also analyze the integrity of the data collectors quantitatively through two different types of analyses namely conflict analysis and learning-based analysis. Our analyses can identify substantial portions of the unreliable data collectors, where learning-based analysis performs better than conflict analysis in performing

the identification task. We examine how far the outcomes of these quantitative analyses and the outcomes of our focused group discussion cross-validate each other in determining the integrity of our data collectors. Considering the outcomes of our focused group discussion as the ground truth, we achieve the best outcomes through a quantitative analysis as precision = 0.47, recall = 0.99, and F1 score = 0.64 in the case of learning-based analysis.

Thus, in this study, we investigate computational techniques to identify potentially unreliable data collectors based on their collected data. Our study presents several new findings covering impacts of religious and socio-geographical aspects on the data collection process (e.g., religious and socio-geographical aspects play a major part in the process of executing tasks by the data collectors), important factors to be considered during scrutiny of the data collectors (e.g., scrutiny over the collected data and collaborative participation of expert reviewers), efficacy of quantitative analyses (e.g., conflict analysis and learning-based analysis) in measuring the integrity of the data collectors, etc.

In the future, we plan to evaluate the efficacy of our approach for evaluating the integrity of data collectors in other mass-scale surveys, where data collectors are engaged to collect data by surveying participants on large scales. We also plan to perform different types of analyses, such as qualitative analysis and quantitative analysis, over our collected survey data. These analyses would be helpful to find out the associations among the problems faced by the pilgrims as well as recommendations proposed by the pilgrims. Finally, we plan to collaborate with Hajj agencies in different countries to employ our survey approach in road to ameliorating the future experiences of the pilgrims.

## **APPENDIX**

### **DETAILS OF THE HAJJ SURVEY QUESTIONNAIRE AND ITS PREPARATION**

Our survey questionnaire has different sections. The sections contain demographic information, communication difficulties, information about getting lost, vehicle shortage, washroom shortage, drinking water shortage, food shortage, problems in Masjid al-Haram, problems in Mina camp, problems in 'Arafah, road directions in Hajj-related places, and other miscellaneous information about the pilgrims during Hajj/Umrah.

The demographic section of our Hajj survey questionnaire collects some personal information of the participant pilgrim such as birth year, gender, the country from where the pilgrim arranged his/her pilgrimage, language, educational qualification, occupation, devices used by the pilgrim for communication during the pilgrimage, year of performing pilgrimage, etc. We use appropriate response types for these questions such as simple text, binary-choice, multiple-choice, etc. Besides, we add the 'Other' option in most of the multiple-choice questions so that participants can provide additional information beyond our listed options. Note that,

during the online survey, we add a binary-choice question (yes/no) on having experience of performing Hajj/Umrah to the demographic section to prevent any accidental response from a person who has not performed pilgrimage yet.

After getting the demographic data from a participant pilgrim, we collect the information about difficulties in verbal communication s/he faced in pilgrimage time using a binary-choice question. If a pilgrim responds 'yes', we further ask whether s/he faces verbal communication difficulties in daily life. This question helps us to know whether the pilgrim has any limitations in communicative skills in real life. Next, we ask the participant pilgrim whether s/he got lost during the pilgrimage or not. If the answer is 'yes', we dig down this problem and collect detailed information about the incident(s) of getting lost such as locations, durations, causes of getting lost, reactions after getting lost, physical conditions after getting lost, etc. After that, we ask the participant pilgrim whether s/he faced any problem in finding vehicles. If s/he responds 'yes', we further ask about the locations of the vehicle inadequacy. Similarly, we ask a pilgrim about the shortage of washrooms, drinking water, and food along with the locations. Next, we ask the participant pilgrim about the problems and possible solutions regarding Hajj/Umrah-related places such as Masjid al-Haram, Mina, and 'Arafah. Note that, questions regarding Mina camp and plain of Arafat do not apply to an Umrah pilgrim, since visiting these places is not mandatory for him/her. Next, we ask about the existing road directions found in the Hajj-related places and whether the available road signs are easy to understand. If the participant pilgrim responds 'no', we collect exact problems that the pilgrim faced in understanding the road signs along with possible solutions. Finally, to know more about other problems, there is an open-ended question asking "What were the other difficulties you faced during Hajj/Umrah?". We summarize different sections of the survey questionnaire in Table 1.

It is worth mentioning that, to learn about real problems and possible solutions regarding Hajj/Umrah-related places, we have some hour-long informal discussions with several pilgrims known to us during the development of our questionnaire. We ask them about the difficulties they faced and possible solutions to minimize these difficulties while performing Hajj/Umrah. As it requires huge time to share own Hajj/Umrah experience, we could not have such a long informal discussion with unknown ones. This happens as an unknown pilgrim naturally does not feel comfortable sharing his/her experience over hour-long discussions, which has also been observed during our survey data collection. Therefore, based on the experiences and suggestions obtained from our known pilgrims as revealed by the long informal discussions, we provide some pre-defined options in our questionnaire. Besides, we enable the 'Other' option in our questionnaire so that we can collect information about new problems and solutions shared by the survey participant pilgrims.

## Survey Form for Hajj Pilgrims

\*Required field

Form Id \_\_\_\_\_

1. Your name (optional) \_\_\_\_\_

2. Your birth year\* \_\_\_\_\_

3. Your gender\* (Please mark only one oval)

Male       Female       Other       Don't want to disclose

4. Your contact number/mobile phone number (optional) \_\_\_\_\_

5. Your email address (optional) \_\_\_\_\_

6. Country you live in\*

Mention the country from where you arranged your Hajj/Umrath pilgrimage \_\_\_\_\_

7. Language(s) you speak\* (Please mention all that apply) \_\_\_\_\_

8. Your highest educational qualification\* (Please mark only one oval)

Below High School/SSC/O-level/Equivalent Degree       High School/SSC/O-level/Equivalent Degree  
 College/HSC/A-level/Equivalent Degree       Above College/Diploma/Equivalent Degree  
 University/Higer Degree

9. Your occupation\* (Please select all that apply)

Student       Teacher       Housewife       IT Professional  
 Doctor/Dentist/Nurse/Pharmacist/Health Professional  
 Defense officer (in Army/Police/similar job)  
 Defense employee (not officer) (in Army/Police/similar job)  
 Government officer       Government employee (not officer)       Private officer  
 Private employee (not officer)       Businessman (large business)  
 Businessman (small business)       Retired       Unemployed       Other: \_\_\_\_\_

10. Which electronic device(s) do you use for communication? (Please select all that apply)

Feature phone (with physical keyboard having no touchscreen)  
 Smartphone/Touchscreen phone       Laptop/Computer/iPad/Tab/Smartwatch  
 None       Other: \_\_\_\_\_

11. Have you ever performed Hajj/Umrath? \* (Please mark only one oval)

Yes       No      Stop filling the form

12. When have you performed Hajj/Umrath? \* (If you have performed multiple times, please separate them by comma. Please mention month and year (e.g., January 2009, October 2015, etc.).)

13. Did you face any difficulty in communicating (talking) with others during Hajj/Umrath? \* (Please mark only one oval)

Yes       No      Skip to question 15

14. Do you face any difficulty in communicating (talking) with others in day to day life? \* (Please mark only one oval)

Yes       No

15. Did you get lost anytime during Hajj/Umrath? \* (Please mark only one oval)

Yes       No      Skip to question 22

16. When did you get lost?\* (Please select all that apply)

Before Tawaf       During Tawaf       After Tawaf  
 Before Sa'i       During Sa'i       After Sa'i  
 While travelling from Mecca to Mina       At the camps in Mina  
 While travelling from Mina to 'Arafah       In the plain of 'Arafah  
 While travelling from 'Arafah to Muzdalifah       While travelling from Muzdalifah to Mina



- During stoning at Jamarat     After stoning at Jamarat     During animal sacrifice  
 While travelling from Mina to Mecca     Cannot remember  
 Other: \_\_\_\_\_

17. When did you get lost for the longest time? \* (If you got lost multiple times, choose the option when you were lost for the longest time. If you got lost once, choose the same option as chosen in the last question.)

- Before Tawaf     During Tawaf     After Tawaf  
 Before Sa'i     During Sa'i     After Sa'i  
 While travelling from Mecca to Mina     At the camps in Mina  
 While travelling from Mina to 'Arafah     In the plain of 'Arafah  
 While travelling from 'Arafah to Muzdalifah     While travelling from Muzdalifah to Mina  
 During stoning at Jamarat     After stoning at Jamarat     During animal sacrifice  
 While travelling from Mina to Mecca     Cannot remember  
 Other: \_\_\_\_\_

18. For how long you were lost? \* (If you got lost multiple times, choose the longest time you were lost for. Please mark only one oval)

- Less than half an hour     1 - 2 hours     2 - 3 hours     3 - 6 hours  
 6 - 12 hours     Less than a day     For 2 - 3 days

19. Which of the following causes were responsible for your getting lost? \* (Please select all that apply)

- Old age     Language barrier  
 Lack of knowledge of necessary names (e.g., hotel name, street name, name of office of pilgrim affairs, etc.)  
 Lack of knowledge of necessary directions (e.g., to hotel, to street, etc.)  
 Leaving the group for a while  
 Entering Masjid al-Haram with one gate and leaving with another  
 Removing the wrist band     All streets look similar  
 Other: \_\_\_\_\_

20. What did you do when you got lost? \* (Please select all that apply)

- I tried to contact local volunteers     I tried to contact agency guides  
 I tried to contact nearby security officers  
 I tried to communicate with my companions/fellow pilgrims over phone  
 I used location sharing services over smartphone  
 I used app-based tracking system over smartphone     I used available map  
 Other: \_\_\_\_\_

21. Did you get sick after/when you got lost? \* (Please mark only one oval)

- Not at all     Partially     Severely

22. Did you get sick during Hajj/Umrab while performing normal activities (not related to getting lost)? \* (Please mark only one oval)

- Not at all     Partially     Severely

23. In case you know - please pick the month(s) of Islamic calendar from the following during which pilgrims are more prone to getting lost. \* (Please select all that apply)

- I don't know  
 Muhamarram     Jumada I     Ramadan  
 Safar     Jumada II     Shawwal  
 Rabi'II I     Rajab     Dhu'l-Qi'dah  
 Rabi'II II     Sha'ban     Dhu'l-Hijjah

24. Did you face any problem in finding vehicles? \* (Please mark only one oval)

- Yes     No    *Skip to question 26*

25. Please mention the location(s) where you faced problems in finding vehicles. \* (Please select all that apply)

- While travelling from Jeddah to Mecca     While travelling from Mecca to Mina  
 While travelling from Mina to 'Arafah     While travelling from 'Arafah to Muzdalifah  
 While travelling from Muzdalifah to Mina     While travelling from Mina to Mecca  
 While travelling from Mecca to Medina     While travelling inside Mecca  
 While travelling inside Medina     Cannot remember  
 Other: \_\_\_\_\_

26. Did you face any problem in finding a washroom/bathroom/toilet? \* (Please mark only one oval)



Yes No[Skip to question 28](#)

27. Please mention the location(s) where you faced problems in finding a washroom/bathroom/toilet.\* (Please select all that apply)

- At hotel in Mecca  At Haram in Mecca  At hotel in Medina  At Haram in Medina  
 Mina  'Arafah  Muzdalifah  Cannot remember  Other: \_\_\_\_\_

28. Did you face any problem in finding a clean/hygienic washroom/bathroom/toilet? \* (Please mark only one oval)

Yes  No [Skip to question 30](#)

29. Please mention the location(s) where you faced problems in finding a clean/hygienic washroom/bathroom/toilet.\* (Please select all that apply)

- At hotel in Mecca  At Haram in Mecca  At hotel in Medina  At Haram in Medina  
 Mina  'Arafah  Muzdalifah  Cannot remember  Other: \_\_\_\_\_

30. Did you face any problem in getting drinking water? \* (Please mark only one oval)

Yes  No [Skip to question 32](#)

31. Please mention the location(s) where you faced problems in getting drinking water.\* (Please select all that apply)

- Mecca  Mina  'Arafah  Muzdalifah  Medina  
 While travelling from Mina to 'Arafah  Cannot remember  Other: \_\_\_\_\_

32. Did you face any problem in getting food? \* (Please mark only one oval)

Yes  No [Skip to question 34](#)

33. Please mention the location(s) where you faced problems in getting food.\* (Please select all that apply)

- Mecca  Mina  'Arafah  Muzdalifah  Medina  
 While travelling from Mina to 'Arafah  Cannot remember  Other: \_\_\_\_\_

34. At which time of a day Masjid al-Haram is crowded the most?\* (Please select all that apply)

- |                                     |                                     |                                    |                                   |                                  |
|-------------------------------------|-------------------------------------|------------------------------------|-----------------------------------|----------------------------------|
| <input type="radio"/> I don't know  | <input type="radio"/> Qiyam al-layl |                                    |                                   |                                  |
| <input type="radio"/> Before Fajr   | <input type="radio"/> After Fajr    | <input type="radio"/> Before Duhr  | <input type="radio"/> After Duhr  | <input type="radio"/> Before Asr |
| <input type="radio"/> After Asr     | <input type="radio"/> Before Magrib | <input type="radio"/> After Magrib | <input type="radio"/> Before Isha | <input type="radio"/> After Isha |
| <input type="radio"/> Jum'ah prayer | <input type="radio"/> Almost always | <input type="radio"/> Never        |                                   |                                  |

35. Did you face any hurdle in entering into Masjid al-Haram? \* (Please mark only one oval)

Yes  No [Skip to question 38](#)

36. What were the hurdles you faced in entering into and moving around Masjid al-Haram? \* (Please select all that apply)

- Overcrowded  Gates were closed  Identifying the gates  
 Paths to Mataf (space surrounding Baitullah) were limited  
 Getting access to Hajar al-Aswad (The Black Stone)  Other: \_\_\_\_\_

37. What are your suggestions to resolve these hurdles you faced in entering into and moving around Masjid al-Haram?\*  
(Please select all that apply)

- Overcrowding should be presented upfront before entering through a gate  
 Pilgrims should be directed towards less crowded gates  
 Entrance to Mataf should be open from all sides  
 Prayer places for female should be designated in Mataf  
 More volunteers should be appointed in Haram  
 Other: \_\_\_\_\_

38. Did you face any hurdle in finding your camp at Mina? \* (Please mark only one oval)

Yes  No [Skip to question 41](#)

39. What were the hurdles you faced in finding your camp at Mina?\* (Please select all that apply)

- The number of volunteers in Mina is limited  
 Communicating (talking) with volunteers in Mina is difficult  
 The tents are identical making it confusing for the pilgrims  
 Finding maps of Mina is difficult  Tracking location from available maps is difficult  
 As I don't use smartphone, I cannot use app-based tracking system  
 I could not take help from surrounding people due to language barrier  
 Other: \_\_\_\_\_

40. What are your suggestions to resolve these hurdles you faced in finding your camp at Mina?\* (Please select all that apply)

- All the tents should bear flag of its country



- All the tents for a country should be held together on a particular road  
 There should be directions on the street guiding to tents of different countries  
 There should be more volunteers in Mina  
 Detailed and comprehensive map of Mina should be provided to the pilgrims  
 Other: \_\_\_\_\_

41. Did you face any hurdle in 'Arafah? \* (Please mark only one oval)

- Yes       No      Skip to question 44

42. What were the hurdles you faced in 'Arafah?\* (Please select all that apply)

- Lighting was not adequate     There was no direction  
 I could not participate in the Jamaat/Congregational Prayer in Nimrah mosque  
 Insufficiency of washroom     There were a lot of insects (mostly ants) in the ground  
 Other: \_\_\_\_\_

43. What are your suggestions to resolve these hurdles you faced in 'Arafah? (optional) (Please select all that apply)

- More lighting should be deployed     Central sound system should be deployed  
 Number of volunteers should be increased     Sanitation facility of toilets should be improved  
 Other: \_\_\_\_\_

44. How frequently did you find directional signs on the roads of Mecca while performing Hajj/Umrath? \* (Please mark only one oval)

- Almost always     Sometimes     Rare  
 Never (Skip to question 49)     I did not think about it / I cannot remember (Skip to question 49)

45. Did you face any difficulty in understanding the existing road direction signs? \* (Please mark only one oval)

- Yes     No      Skip to question 49

46. What were the difficulties you faced in understanding road directions?\* (Please select all that apply)

- As the directions were in Arabic/English, I could not understand them  
 The direction signs were too small     The direction signs were not properly visible at night  
 Other: \_\_\_\_\_

47. What are your suggestions to make road directions easy to understand? (optional) (Please select all that apply)

- Symbolic language (e.g., a red forearm means STOP) usage should be increased with road directions  
 Adding images of next place with an arrow as self-explaining symbols  
 Conducting campaigns to increase awareness of pilgrims  
 Other: \_\_\_\_\_

48. What were the other difficulties you faced during Hajj/Umrath? (optional) (Please select all that apply)

- Due to scorching heat at 'Arafah, I got sick (e.g., asthma, hypertension, heat stroke, etc.)  
 I got separated from my male companion(s) for a while  
 I lost my baggage during Hajj/Umrath  
 It was difficult to contact my family at home  
 I could not avail Wi-Fi service at the hotels  
 It was difficult to find transportation during 8th-13th Dhu'l-Hijjah  
 Local people were not very cooperative  
 Losing necessary documents (e.g., passport, NID, etc.)  
 I lost my baggage when I was returning from Hajj/Umrath  
 I experienced theft  
 Separate washrooms for females were unavailable  
 Other: \_\_\_\_\_

49. What are your suggestions to eradicate these hurdles?(optional) (Please select all that apply)

- Conducting campaigns to increase awareness of pilgrims  
 Separate washrooms should be arranged for women  
 While taking bus, "Female first" policy should not be followed as it causes more trouble than convenience. Rather "family together" policy might be adopted  
 Medical camp should be increased and made easily accessible  
 There should be more volunteers who can communicate in English/other languages  
 Other: \_\_\_\_\_

## THANK YOU



## REFERENCES

- [1] A. Wiggins and Y. He, "Community-based data validation practices in citizen science," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput.*, Feb. 2016, pp. 1548–1559.
- [2] D. Hellmann, C. Maitland, and A. Tapia, "Collaborative analytics and brokering in digital humanitarian response," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput.*, Feb. 2016, pp. 1284–1294.
- [3] A. Nugraha, I. A. Wahono, and T. Inoue, "An initial study of collaborative refinement of student-created questions," in *Proc. Conf. Companion Comput. Supported Cooperat. Work Social Comput.* New York, NY, USA: ACM, Oct. 2020, pp. 359–363, doi: [10.1145/3406865.3418327](https://doi.org/10.1145/3406865.3418327).
- [4] M. Dittus, G. Quattrone, and L. Capra, "Analysing volunteer engagement in humanitarian mapping: Building contributor communities at large scale," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput.*, Feb. 2016, pp. 108–118.
- [5] M. Kosmala, A. Wiggins, A. Swanson, and B. Simmons, "Assessing data quality in citizen science," *Frontiers Ecol. Environ.*, vol. 14, no. 10, pp. 551–560, Dec. 2016.
- [6] R. M. Borromeo and M. Toyama, "An investigation of unpaid crowdsourcing," *Hum.-Centric Comput. Inf. Sci.*, vol. 6, no. 1, pp. 1–19, 2016.
- [7] A. Voida, Z. Yao, and M. Korn, "(Infra) structures of volunteering," in *Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Social Comput.*, Feb. 2015, pp. 1704–1716.
- [8] M. Bernstein, M. Bright, E. Cutrell, S. Dow, E. Gerber, A. Jain, and A. Kulkarni, "Micro-volunteering: Helping the helpers in development," in *Proc. Conf. Comput. Supported Cooperat. Work Companion (CSCW)*, 2013, pp. 85–88.
- [9] C. Kiene, K. Shores, E. Chandrasekharan, S. Jhaver, J. Jiang, B. Dym, J. Seering, S. Gilbert, K. Lo, D. Y. Wohn, and B. Dosono, "Volunteer work: Mapping the future of moderation research," in *Proc. Conf. Companion Publication Comput. Supported Cooperat. Work Social Comput.*, Nov. 2019, pp. 492–497.
- [10] R. Joshi, "Evaluating the impact of human activities during the Mahakumbh 2010 fair on elephants in the Shivalik elephant reserve," *Tropical Natural History*, vol. 13, no. 2, pp. 107–129, 2013.
- [11] GASTAT. (2019). *Hajj Statistics 2019*. Accessed: Jun. 8, 2021. [Online]. Available: [https://www.stats.gov.sa/sites/default/files/haj\\_40\\_en.pdf](https://www.stats.gov.sa/sites/default/files/haj_40_en.pdf)
- [12] H. Taibah, S. Arlikatti, and S. Andrew, "Risk communication for religious crowds: Preferences of Hajj pilgrims," *Disaster Prevention Manage.*, vol. 27, no. 1, pp. 102–114, Feb. 2018.
- [13] J. Fourati, B. Issaoui, and K. Zidi, "Literature review of crowd management: A Hajj case study," in *Proc. ICINCO*, 2017, pp. 346–351.
- [14] A. Shujaa and S. Alhamid, "Health response to Hajj mass gathering from emergency perspective, narrative review," *Turkish J. Emergency Med.*, vol. 15, no. 4, pp. 172–176, 2015.
- [15] Y. A. Alaska, A. D. Aldawas, N. A. Aljerian, Z. A. Memish, and S. Suner, "The impact of crowd control measures on the occurrence of stampedes during Mass Gatherings: The Hajj experience," *Travel Med. Infectious Disease*, vol. 15, pp. 67–70, Jan./Feb. 2017.
- [16] M. Hoseinpourfard, M. Ghanei, S. Tofighi, and M. Izadi, "The emergence of Hajj stampedes: Lessons for Hajj trauma centers," *Trauma Monthly*, vol. 22, no. 4, Jul. 2017, Art. no. e39455.
- [17] M. Ganjeh and B. Einollahi, "Mass fatalities in Hajj in 2015," *Trauma Monthly*, vol. 21, no. 5, Nov. 2016, Art. no. e43253. <https://pubmed.ncbi.nlm.nih.gov/28184365>
- [18] (Sep. 2015). *More Than 700 Pilgrims Die in Crush in Worst HAJ Disaster for 25 Years*. Reuters. [Online]. Available: <https://www.reuters.com/article/us-saudi-hajj-casualties-idUSKCN0ROORN20150924>
- [19] N. Reeves, R. Tinati, S. Zerr, M. G. Van Kleek, and E. Simperl, "From crowd to community: A survey of online community features in citizen science projects," in *Proc. ACM Conf. Comput. Supported Cooperat. Work Social Comput.*, Feb. 2017, pp. 2137–2152.
- [20] R. P. Karumur, B. Yu, H. Zhu, and J. A. Konstan, "Content is king, leadership lags: Effects of prior experience on newcomer retention and productivity in online production groups," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–13.
- [21] R. P. Karumur, T. T. Nguyen, and J. A. Konstan, "Early activity diversity: Assessing newcomer retention from first-session activity," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput.*, Feb. 2016, pp. 595–608.
- [22] H. M. Almuhrizi and A. M. Alsawafi, "Muslim perspectives on spiritual and religious travel beyond Hajj: Toward understanding motivations for Umrah travel in Oman," *Tourism Manage. Perspect.*, vol. 24, pp. 235–242, Oct. 2017.
- [23] H. Alnabulsi, J. Drury, V. L. Vignoles, and S. Oogink, "Understanding the impact of the Hajj: Explaining experiences of self-change at a religious mass gathering," *Eur. J. Social Psychol.*, vol. 50, no. 2, pp. 292–308, 2020.
- [24] A. M. Alkhoubi, "Wayfinding for pedestrians in the crowded areas of Al-Hajj: How can wayfinding system designs increase the efficiency of wayfinding and navigation performances for pedestrian pilgrims during the Islamic pilgrimage (Al-Hajj)," Ph.D. dissertation, Univ. Minnesota, Barbara Martinson, Minneapolis, MN, USA, Apr. 2020.
- [25] E. A. Felemban, F. U. Rehman, S. A. A. Biabani, A. Ahmad, A. Naseer, A. R. M. A. Majid, O. K. Hussain, A. M. Qamar, R. Falemban, and F. Zanjir, "Digital revolution for Hajj crowd management: A technology survey," *IEEE Access*, vol. 8, pp. 208583–208609, 2020.
- [26] E. Aflalo, "Students generating questions as a way of learning," *Act. Learn. Higher Educ.*, vol. 22, no. 1, pp. 63–75, Mar. 2021, doi: [10.1177/1469787418769120](https://doi.org/10.1177/1469787418769120).
- [27] J. Bardram and A. Doryab, "Activity analysis: Applying activity theory to analyze complex work in hospitals," in *Proc. ACM Conf. Comput. Supported Cooperat. Work (CSCW)*, 2011, pp. 455–464.
- [28] G. Hsieh and R. Kocielnik, "You get who you pay for: The impact of incentives on participation bias," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput.*, Feb. 2016, pp. 823–835.
- [29] C. G. Harris, "The effects of pay-to-quit incentives on crowdworker task quality," in *Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Social Comput.*, Feb. 2015, pp. 1801–1812.
- [30] B. Knowles, M. Rouncefield, M. Harding, N. Davies, L. Blair, J. Hannon, J. Walden, and D. Wang, "Models and patterns of trust," in *Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Social Comput.*, Feb. 2015, pp. 328–338.
- [31] L. P. Robert, "Monitoring and trust in virtual teams," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput.*, Feb. 2016, pp. 245–259.
- [32] Y. Wang and D. Redmiles, "The diffusion of trust and cooperation in teams with individuals' variations on baseline trust," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput.*, Feb. 2016, pp. 303–318.
- [33] W. Huang, T. Lu, H. Zhu, G. Li, and N. Gu, "Effectiveness of conflict management strategies in peer review process of online collaboration projects," in *Proc. 19th ACM Conf. Comput.-Supported Cooperat. Work Social Comput.* New York, NY, USA: ACM, Feb. 2016, pp. 717–728.
- [34] P. Singh and B. Suri, "Quality assessment of data using statistical and machine learning methods," in *Computational Intelligence in Data Mining*, vol. 2, L. C. Jain, H. S. Behera, J. K. Mandal, and D. P. Mohapatra, Eds. New Delhi, India: Springer, 2015, pp. 89–97.
- [35] L. Ehrlinger, V. Haunschmid, D. Palazzini, and C. Lettner, "A DaQL to monitor data quality in machine learning applications," in *Database and Expert Systems Applications*, S. Hartmann, J. Küng, S. Chakravarthy, G. Anderst-Kotsis, A. M. Tjoa, and I. Khalil, Eds. Cham, Switzerland: Springer, 2019, pp. 227–237.
- [36] N. U. Okafor, Y. Alghorani, and D. T. Delaney, "Improving data quality of low-cost IoT sensors in environmental monitoring networks using data fusion and machine learning approach," *ICT Exp.*, vol. 6, no. 3, pp. 220–228, Sep. 2020. [Online]. Available: <https://www.sciedirect.com/science/article/pii/S2405959520300862>
- [37] G. F. Kuder and M. W. Richardson, "The theory of the estimation of test reliability," *Psychometrika*, vol. 2, no. 3, pp. 151–160, Sep. 1937.
- [38] W. Xie, W. A. Bainbridge, S. K. Inati, C. I. Baker, and K. A. Zaghloul, "Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe," *Nature Hum. Behav.*, vol. 4, no. 9, pp. 937–948, Sep. 2020.
- [39] C. Damsa, M. Langford, D. Uehara, and R. Scherer, "Teachers' agency and online education in times of crisis," *Comput. Hum. Behav.*, vol. 121, Aug. 2021, Art. no. 106793.
- [40] J. Kehoe, "Basic item analysis for multiple-choice tests," *Practical Assessment, Res., Eval.*, vol. 4, no. 1, p. 10, 1994.
- [41] Statista. (2021). *Internet Penetration Rate in Bangladesh From 2010 to 2019*. Accessed: Jun. 8, 2021. [Online]. Available: <https://www.statista.com/statistics/764102/internet-penetration-rate-bangladesh/>
- [42] Zoom. (2011). *Video Conferencing, Web Conferencing, Webinars, Screen Sharing*. Accessed: Jun. 8, 2021. [Online]. Available: <https://zoom.us/>
- [43] J. Alsakran, X. Huang, Y. Zhao, J. Yang, and K. Fast, "Using entropy-related measures in categorical data visualization," in *Proc. IEEE Pacific Visualizat. Symp.*, Mar. 2014, pp. 81–88.
- [44] D. Heerwagh and G. Loosveldt, "Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality," *Public Opinion Quart.*, vol. 72, no. 5, pp. 836–846, Dec. 2008.

- [45] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982.



**SAIFUL ISLAM SALIM** received the B.Sc. degree from the Department of CSE, Bangladesh University of Engineering and Technology (BUET), where he is currently pursuing the M.Sc. degree in computer science. He is currently working as a Graduate Research Assistant with the Department of CSE, BUET, under the supervision of Prof. A. B. M. Alim Al Islam. His research interests include machine learning, security, human-computer interaction, image processing, multi-objective optimization, intelligent transportation, and robotics.

**NAJLA ABDULRAHMAN AL-NABHAN** received the B.S. (Hons.) and master's degrees in computer science from King Saud University, and the Ph.D. degree in computer science from the College of Computer and Information Science (CCIS), King Saud University. She was a Visiting Ph.D. Student with George Washington University (2012–2013). She is currently working as an Assistant Professor in computer science with the CCIS, King Saud University. Her research interests include the Internet of Things, emergency and crowd management systems, wireless sensor networks, mobile computing, distributed systems, cloud and fog computing, smart grid, and network security.



**MASFIQUR RAHAMAN** received the B.Sc. (Engg.) degree from the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET). He is currently working as a Graduate Research Assistant with the Department of CSE, BUET, under the supervision of Prof. A. B. M. Alim Al Islam. His research interests include machine learning, human-computer interaction, ubiquitous computing, structural health monitoring, and intelligent transportation.

**NAFISA ISLAM** received the B.Sc. and M.Sc. degrees from Bangladesh University of Engineering and Technology (BUET), in 2017 and 2021, respectively. She held several academic and research positions with Sonargaon University, Stamford University Bangladesh, United International University, and Bangladesh University of Engineering and Technology. In 2021, she joined as a Lecturer with the Department of Civil and Environmental Engineering, Islamic University of Technology (IUT). Her research interests include air pollution monitoring with particular focus on meteorological influence, heavy metal emission profile development, and data analysis and interpretation.



**TARIK REZA TOHA** received the B.Sc. degree in computer science from BUET, in 2017, where he is currently pursuing the M.Sc. degree in computer science. He has been working as a Lecturer with the Department of Textile Machinery Design and Maintenance (TMMD), Bangladesh University of Textiles (BUTEX), since December 2018. Previously, he has worked as a Research Assistant with the Department of Computer Science and Engineering (CSE), Bangladesh University of Engineering and Technology (BUET), under the supervision of Prof. A. B. M. Alim Al Islam. His research interests include computer vision, deep learning, embedded systems, and green computing.

**JANNATUN NOOR** received the B.Sc. and M.Sc. degrees in computer science from BUET. She has been working as a Lecturer with the Department of CSE, BRAC University, since September 2018. Besides, she is working

as a Graduate Research Assistant under the supervision of Prof Dr. A. B. M. Alim Al Islam with the Department of CSE, Bangladesh University of Engineering and Technology (BUET). Prior to BRACU, she worked as a Team Lead of the IPV-Cloud Team, IPvision Canada Inc. Her research interests include cloud computing, wireless networking, data mining, big data analysis, and the Internet of Things.

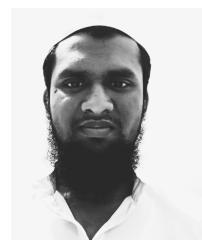


**ADNAN QUAIUM** received the B.Sc. (Engg.) degree in electrical and electronic engineering from the Department of Electrical and Electronic Engineering, Islamic University of Technology (IUT), Dhaka, Bangladesh, and the M.Sc. (Engg.) degree in electrical engineering from the Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology (TU Delft), Delft, The Netherlands. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka. He is currently working as an Assistant Professor with the Ahsanullah University of Science and Technology, Dhaka. His research interests include environmental monitoring and sensing, networking and systems, machine learning, and HCI.

**AAIYEESHA MOSTAK** received the B.Sc. degree from the Department of CSE, BUET, where she is currently pursuing the M.Sc. degree in computer science, under the supervision of Prof. A. B. M. Alim Al Islam. Her research interests include neural computing, machine learning, human-computer interaction, and the Internet of Things.



**MAINUL HOSSAIN** received the B.Sc. (Engg.) degree from the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET), where he is currently pursuing the M.Sc. degree in computer science. He is currently working as a Research Assistant with the Department of CSE, BUET, under the supervision of Prof. A. B. M. Alim Al Islam. His research interests include machine learning, human-computer interaction, and the Internet of Things.



**MD. MASUM MUSHFIQ** received the B.Sc. degree in computer science from BUET, in 2019, where he is currently pursuing the M.Sc. degree in computer science. He is currently working as a Lecturer with the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET). His research interests include computer security, modeling and simulation, and intelligent transportation systems.



**A. B. M. ALIM AL ISLAM** received the B.Sc. (Engg.) and M.Sc. (Engg.) degrees in computer science and engineering from the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, and the Ph.D. degree in computer engineering from the School of ECE, Purdue University, USA, in 2012, with a prestigious fellowship. He is currently working as a Professor with the Department of Computer Science and Engineering, BUET. His research interests include human-computer interaction, ubiquitous computing, networking and systems, security, modeling and simulation, intelligent transportation, environmental monitoring and sensing, and reliability analysis. He has coauthored more than 100 journals and conference papers. He has presented his research papers in different well-reputed international conferences. He delivered speeches in different national and international venues. He achieved several research awards covering best paper awards, best poster awards, best demo awards, research grants, and travel grants.