Call for Service: Characterizing and Modeling Police Response to Serviceable Requests on Facebook

Niharika Sachdeva
IIIT-Delhi, India
niharikas@iiitd.ac.in

Ponnurangam Kumaraguru
IIIT-Delhi, india
pk@iiitd.ac.in

ABSTRACT

Social media platforms have obtained substantial interest of police to connect with residents. This has encouraged residents to report day-to-day law and order concerns such as traffic congestion, missing people, and harassment by cops on these platforms. In this paper, we study day-to-day concerns shared by residents on social media and police response to such concerns. Based on the input of police experts, we define concerns that require police response and attention, as a serviceable request. We provide insights on six textual attributes that can identify serviceable posts. We find such posts are marked by high negative emotions, more factual, and objective content such as location and time of incidences. We show that police response time varies depending upon the kind of serviceable requests. Our work explores a series of statistical models to predict serviceable posts and its different types. We conclude the paper, discussing the implication of our findings on police practices and design needs for possible technological interventions. These technological interventions will help increase the interactions between police and residents and thereby increasing the well-being and safety of society.

Author Keywords

Social media; police; citizens; service; measures

ACM Classification Keywords

H.5.3 Group and Organization Interfaces: Web-based Interaction

INTRODUCTION

Law and order concerns are one of the major disquiets of urban societies in day-to-day life. These concerns often bear detrimental effects on the psychological well-being of the residents and society at large [32]. Police are one of the most omnipresent and ubiquitous bodies of society that addresses resident's well-being, concerns and requests [27]. Responsiveness of the police officers to these requests in terms of promptness and action they take determines the overall satisfaction of community and police accountability [22, 36]. To improve police accountability, modern police departments explore innovative mechanisms and technologies to keep themselves available and connected with residents [31].

In recent years, due to the enormous and pervasive reach of social media, police has realized it as a prominent medium to support interaction with residents [12, 25]. Online interactions through social media can influence residents' perceptions about safety, police efforts, and the relationship between them [60]. Research has examined the efficacy of social media in a diverse set of scenarios like crises (natural and manmade) and socio-political upheavals [6, 43, 49]. In such scenarios, exchanges between residents, emergency responders, and organizations for operational collaboration have also been explored [50, 58]. Research also exhibits the role and consequences of social media use by police to provide support during crises [5, 12, 25, 34]. Despite its usefulness in crises, social media role to enable police in responding to day-to-day concerns of residents remains largely unexplored.

In this research, we examine resident requests concerning day-to-day issues that elicit police response on social media. Requests to which police should respond, evaluate or take action are considered as *serviceable requests* [55]. Owing to the massive volume of content on social media, identifying serviceable requests from general conversation and spam messages results in additional load on already resource constrained police organizations [23, 30]. To the best of our knowledge, this is the first work that determines the possibility of identifying *what type of posts should get a response from police (serviceable requests)* and estimates how long police would take to respond a particular type of service request. To this end, we formulate our specific research goals as:

- Investigating attributes that can characterize and identify residents' service requests and expected police response to such posts;
- Identifying patterns and estimating the response time by the police to residents.

For the above-stated goals, we study resident generated content from 85 official (and public) Facebook pages of police organizations in India. We contacted the state police organizations to understand what posts police officers can service on social media. We make following contributions: a) based on the response elicited by police for the residents posts, we develop a hybrid methodology that integrates text characteristics and annotations from expert police officers to identify different serviceable posts; b) create a new police experts an-

CSCW '17, February 25-March 01, 2017, Portland, OR, USA

DOI: http://dx.doi.org/10.1145/2998181.2998292

notated dataset to help identify attributes of serviceable requests; c) propose six text-based measures – emotions, cognitive and interpersonal expression, low-level linguistic attributes, question-centric approach, topic-based, and entity based attributes to characterize posts that qualify to be serviceable. Using this approach, we are able to predict with 87% accuracy if a post is serviceable and further predict its different types with an accuracy of 65%; and d) show that police response time varies significantly with different serviceable requests. Using robust survival analysis, we estimate the expected time when residents could get a reply from police. Our observations show a decrease in response time associated with the serviceable requests that can be immediately resolved, thus suggesting that successful identification of serviceable posts may ultimately result in systems that facilitate in extending timely police support to residents and improve police responsiveness towards residents.

Our work contributes and builds upon the prior knowledge in the CSCW community on using social media to analyze human behavior for safety collaboration [6, 50, 58]. This work has a two-fold contribution to Computer-Mediated-Communication; on one hand, it can help in data-driven policing by identifying serviceable requests from a large volume of posts that should elicit police response and on the other hand, can help residents to draft a post such that it gets a response from police.

BACKGROUND AND RELATED WORK

Aiding police and citizens through social media has gained the significant interest of researchers working on Computer-Mediated-Communication (CMC) [25]. The focus of CMC researchers in this domain has been on studying varied aspects such as communication needs, social media response strategies, and grassroots coordination and support [12, 25]. With few exceptions, these studies analyze large-scale crisis situation such as earthquake, riot, fire, and floods [6, 12, 43, 44, 49]. In [12, 21], authors discussed effective communication strategies for police organization to provide timely information to citizens during crises. Studies also investigate communication patterns of different police departments and styles of engagement during crises [12]. Further research also explores communication challenges arising due to diversity in groups that seek police assistance through social media especially during safety critical situations [34].

Emphasizing social media's role in crisis informatics, various studies highlight that citizens can become a valuable resource for gathering crisis information [42, 51, 56]. Palen et al. have highlighted public reactions during critical events such as societal upheavals [24]. Prior work also investigates the possibility of developing real-time systems to extract information for first responders such as emergency response teams [5, 20, 38, 51, 57]. These studies highlight information that citizens share during a crisis and how it is utilized. A major challenge in such studies has been monitoring a large volume of public information being shared by residents to identify information of interest [23, 28]. Most of these studies characterize textual content shared during crisis events [56, 23, 28]. It remains, however, unclear if this vast understanding of content characteristics available for event-driven situations can be helpful for police to understand characteristics of a resident's request for service in the context of routine day-to-day policing.

Few studies investigate the differences between behavioral attributes of residents and police during day-to-day communication [18, 46, 47, 48]. These studies highlight how communication and types of information are exchanged in discussions between police and residents. Research also shows that a large number of posts on social media for police do not receive any response [48]. However, complementary work on the responsiveness of the police officers to the resident's request has been extensively studied in criminology and social science research [22, 36]. Key contribution of these criminology studies are: a) how well police deals with and responds to these service requests defines the overall satisfaction of community with the police and b) these requests account for the major information source contributing towards collective action for improved safety and psychological well-being of residents. We note that characterization and identification of how serviceable residents' posts differ from general conversations (hereafter referred as non-serviceable requests) between police and residents on social media remain largely unexplored. In this work, we pursue the use of social media as a source of crowd generated requests and a means to characterize service requests that need response from the police organization. We believe this understanding can help improve responsiveness of police to day-to-day requests by residents highlighting the posts that need their action and response.

Our work relates to studies on police strategies to use and adopt social media for improved engagement with residents [28, 39]. These studies highlight challenges for police organization such as necessary formal approval from higher authority to engage in public conversations, insufficient tools and training to use social media, and limited resources to respond on social media [7, 23, 30]. Response from police makes other residents in the social network realize that police hold themselves accountable to residents' requests [12, 48]. Despite the importance of understanding response patterns, we find that such patterns during routine interactions on social media remain largely unexplored.

Our research is uniquely placed at the intersection of the vast literature on social media informatics using textual extraction and forecasting, interactions through CMC for police organizations. We believe that investigating mechanisms for extracting serviceable request from a large volume of content generated during day-to-day interactions between police and residents on social media is a need for several reasons. First, serviceable requests going unnoticed by the police organization may have significant influence on the assistance seeking patterns of residents [12, 48]; multiple such incidences may have implications for future interactions and may influence other resident's opinion about satisfaction with police [29]. Second, understanding serviceable requests may help in creating intelligence regarding current crime landscape and necessary reforms needed. Thus to learn about attributes that prompt serviceable requests, we first characterize the structure of textual posts shared with police on Facebook. We follow this up with creating machine-learning models that can identify and classify serviceable posts. Finally, we look at how police response varies to these requests to understand implication of this communication.

The context - Indian Policing

India is second largest populous country employing the second largest police force in the world [13, 53]. For every 100,000 residents in India, 130 police officers are deputed [53]. This number is far less than the number suggested by UN guideline i.e. 270-280 police personnel per 100,000 residents [16]. To alleviate this shortage, police adopts multiple media to stay connected with the community. Citizens can individually connect with the police by calling the telephonic helpline (100 similar to 911), writing to official email, or visiting a police station [26]. Realizing the importance of community involvement, police has introduced provisions for community members to watch their neighborhood, join beat constables to perform beats, and arrange get-togethers to understand neighborhoods [26]. Owing to the popularity of social media such as Facebook, Indian Police is adopting and encouraging residents to use social media [41]. Many cities have set up their Facebook pages under the supervision of senior officials to answer citizen's concerns. More than 3 Million people have liked these pages. Police have issued guidelines on how residents can use Facebook to report incidences and ask to include information such as time, date and place of the incidence while reporting. Police also issue action taken reports and conducts these social media endeavors under the guidelines of the Govt. of India [14].

Research Questions

In view of the above literature and our objective to characterize attributes of serviceable requests from residents to police, we address the following research questions:

- *RQ 1:* What are the attributes that differentiate – a) serviceable and non-serviceable requests and b) sub-types of serviceable requests on social media w.r.t content characteristics such as linguistic and emotional attributes of a post?
- *RQ 2:* How does police response time vary between serviceable and non-serviceable posts made on social media?
- *RQ 3:* Can machine learning approaches be used to identify serviceable posts? If yes, can we further classify them into different sub-types of the serviceable requests using post characteristics (content and metadata)?

DATA AND METHODS

Our interest in studying serviceable posts is to investigate the resident's law and order related posts that should elicit police response. To provide a comprehensive understanding, we adopt a mixed method approach, analyzing the posts both qualitatively and quantitatively on two dimensions: content attributes and police response time. This analysis helps us understand the nature of the citizen's concern and response time from police. Finally, based on the understanding of the dimensions mentioned above, we develop prediction and statistical models to achieve our goal i.e. identify serviceable posts.

Collection Methodology

Our study utilized data from a number of public and official Facebook pages of police departments in India. We employed a variety of mechanisms to identify these pages and then to filter posts and comments for analysis. We started by identifying all the police departments in India on Facebook; for this purpose, we referred to a government website [1] that provides a list of all police departments. Using our initial list, we were able to recognize 100 police departments on Facebook. Next, we verified these department Facebook pages for their authenticity and credibility. We manually checked if the pages were linked to the official police department's webpage (e.g. http://www.bcp.gov.in/). After this cleaning, we were left with 85 police departments' pages on Facebook. In the 1000 annotated posts, we find that only four posts were in regional languages. English is the common business language used in India that can be a reason for the majority of the posts being in English in our dataset.

We then collected data using the Facebook Graph API [2] for four months (Aug - November 2015). We obtained all posts (wall posts, or content posted by residents on police page; and status updates, or content posted by police on its page) from these pages. To study serviceable requests sent by residents to police on social media, we filtered the posts that residents posted in the four months, giving us 22,213 wall posts. Further, to evaluate how many posts received a response on these pages, we collected all comments and 'likes' associated with the posts, including their time of creation. Our data did not include private messages that people might have sent to police using Facebook, or other forms of "spill-over" activity that may be present in users' personal Facebook profiles.

Data Sampling & Police Officers (Human Expert) Coding

After collecting the dataset, we randomly selected 1,000 posts for expert annotators. As we were interested in identifying serviceable posts relevant to police, we approached a state police department's social media team to annotate the posts. Six police officers of the group performed the annotation tasks. Police officers were provided a link to an annotation portal which we developed, where they could provide their responses. Officers were shown the text, photograph, video and description or story (if any) associated with the post. Each officer labeled the wall post as serviceable or non-serviceable using the coding scheme suggested in the literature, which defines serviceable request as *a message that solicits a response in a form of an action or information from the police* [55].

If a police officer marked a post as serviceable, they further labeled it with the action needed on that post i.e. forwarding the post to the concerned police station, need more information from the resident, give a solution to the query asked, reply thanks to the post of the residents. This coding scheme was developed by [48]. Two police officers annotated each post; we determined the final label for posts based on the mutual vote (where both the annotators agreed). We also found that many of these posts did not have any text (message, story or description) associated with them; these posts were mostly

[1] http://arunpol.nic.in/

[2] https://developers.facebook.com/docs/graph-api

videos and photographs displaying some incident. For the purpose of this study, we were interested in using the content of the post to identify serviceable posts; we excluded such posts (with no text) leaving us with 663 expert annotated posts. Among these, 300 posts received a reply from the police. Next, we calculated Fleiss Kappa to assess the reliability of the labels generated after the annotation tasks. We found that inter-annotator reliability was 0.77 showing substantial agreement among the police officers and was comparable to similar tasks performed on social media data in the literature.

Terminology and Data Categorization

We now explain the terminology used in the paper to categorize types of posts prevalent in our dataset. We categorize the post that should get a response from police as serviceable requests and based on expected police action suggested in literature [48], these are categorized into four sub-types:

(a) **Forward**: This sub-type category included posts which had enough information and could be forwarded to appropriate authorities for action. For instance, a resident posted, *Date : 4/11/2015 (Wednesday), Time : 10:17 pm, Number : [withheld], Location : [withheld], Violations : Crossing line by way too much obstructing the vehicles which were coming from [withheld] entrance later he jumped the signal ....*

(b) **Give Solution**: These posts mostly included queries by residents to police that could be answered without any detail; resident asks, *Admin !! Can U Explain to Me How Two Challans On Same Date Same Time in Just 5 Minutes Gap !! How Its Possible ?? Any Thing Wrong ??*

(c) **Acknowledge** with thanks: Acknowledged posts sub-type consisted of the posts to which the police wrote "thanks for sharing the information" or "thanks for the appreciation." For instance, resident remarks, *Chennai City Traffic Police a humble salute from a fellow Chennaiite for the commendable job in such rains!!*

(d) **Need more details**: In these resident's posts, police inquired more details so that action could be taken, e.g., a resident asks, *Cops driving wrong side [of road] near XXX hotel .. what action will be taken against them ?* This post lacks information such as time and date when the incident happened.

Table 1 shows the number of posts, likes, and comments on residents' post in our annotated dataset for each category.

Post Type	Posts	Likes	Comments
Serviceable			
Forward	286	1383	661
Give solution	88	183	121
Thanks	72	1288	63
Need more information	104	1245	258
Total	**550**	**4099**	**1103**
Non-Serviceable			
Total	**113**	**316**	**32**

Table 1: Number of posts in the annotated dataset that residents made on Facebook for police action and reply.

MEASURES AND FEATURE SET

Residents often use different language styles in posts while expressing their concerns and asking queries to police. We begin by discussing various handcrafted features used to characterize different language styles in serviceable posts. Next, we complement the handcrafted features with LDA and NMF based features that help automatically discover the latent dimensions and induce semantic features in our data as handcrafted features may not account for all patterns in the data. In this section, we explain different content-based features to detect serviceable posts.

Emotional Attributes: Individual safety and threat are associated with a high emotional activity of individuals [11]. Emotions, being a strong tool for expression of thoughts and opinions, influence expression of attitude and behavior of individuals. To estimate emotional expression in the residents' requests, we consider three measures: a) emotional state, b) emotional intensity, and c) emotional Valence. Emotional states are measured in terms of "anger", "disgust", "fear", "joy", and "sadness" expressed in the residents' posts using IBM Watson's Alchemy API. [3] Research validates use of Alchemy API and finds high correlation between emotional states and sentiment extracted using Alchemy API [3, 61]. The results obtained from this API are reported to be better than state-of-the-art models on social media data [35].

We measure emotional intensity based on the psychological arousal measures of words given in the ANEW dictionary (Affective Norms for English Words). [4] Emotional valence is measured in terms of Positive Affect, Negative Affect, Anxiety – we make use of the popular and well-validated psycholinguistic tool LIWC [5]. Prior social media analytics work uses and validates applicability of these measures on social media data [5, 9, 45].

Cognitive and Interpersonal Attributes: Individuals encountering law and order situations may either be directly or indirectly affected. When directly affected, residents talk about their own experiences whereas if indirectly affected, they focus on others; such experiences are marked with high cognitive uncertainty [32]. Hence, we study some forms of cognitive uncertainty, interpersonal focus, and cognitive orientation mentioned in the posts by residents using LIWC; prior literature has examined these in social media content [10, 9]: 1) measures of *interpersonal focus* given by pronoun use, such as 1st person singular, 1st person plural, 2nd person, 3rd person singular, 3rd person plural, and impersonal pronouns; 2) measures of *cognition* given by the categories "cognitive mech", "tentativeness" and "discrepancy".

Linguistic Attributes: Subject to the social and psychological ecosystem, linguistic styles can assess the resident's use of language and identify evidence about their behavioral characteristics [8]. To measure the variation in linguistic attributes, we consider four measures: 1) Objectivity, 2) Tenses, 3) Lexical Density and 4) Parts-Of-Speech. Posts containing

[3]http://www.alchemyapi.com/api/emotion-analysis
[4]http://www.csc.ncsu.edu/faculty/healey/maa-16/text/
[5]http://liwc.wpengine.com/

situational information about law and order contain more objective than subjective words [56]. We use OpinionFinder subjectivity lexicon to count the number of subjective and objective tags in a post. [6] We capture various other Linguistic Attributes using LIWC: 1) measure of "past", "present", and "future" tenses. This can help quantify residents' complaints about events in the past and concerns that may cause law and order problems in future; 2) measure of *Lexical Density* given by use of "article", "verb", "auxverb", "adverb","preposition", "negation","quantifier", and "number".

Research shows that Parts-Of-Speech are associated with emotional states and information given in the posts [59]. For instance, information is associated with Proper nouns such as "MG Road" Police Station and emotion is associated with adjective such as "disgusted" resident. We used Stanford's Part-of-speech (POS) tagger to label words in each post. [7]

Question Asking Attributes: We measured information seeking behavior to identify serviceable posts and its subtypes. For this purpose, we captured various posts containing words "who", "how", "why", "what", "where", "whom" and containing a "?" as well. Since all questions could not be captured using this method, we made heuristic rules e.g. "Can you..." and "Do you want" to capture all questions that were part of the posts.

Entity-Based Attributes: While reporting a violation, police recommends a format for residents, including, location, date and time, and the name of the affected individual / organization in their posts. Thus indicating a correlation between the presence of these entities and serviceable posts. We tagged each post with various entities such as "people", "companies", "organizations", "cities", "geographic features", "facility" using Alchemy API to identify entities. [8] Next, we also used heuristic based rules and a regular expression to capture if a post contains date and time.

Topical Attributes: Generating intelligence from the content shared on social media requires capacity to cluster and link related concepts [17]. We employ n-gram (unigrams, bigrams, and trigram) analysis on the content posted. We adopt LDA topic modeling to examine the nature of topics discussed. We first train these two models on annotated posts from our dataset. Next, we derive an optimal number of top topics (K) in these posts, then, we iteratively apply this method with different values of k, chosen in an empirical data-driven manner. We obtain optimal value of k via minimizing K-L divergence and maximizing log-likelihood, which is found to be 7; following this, we got two annotators label these seven topics. Topics derived for these models include: fines issued for violations, traffic light violations, traffic congestion in different areas, etc. We summarize these topics in the Table 2.

Non-negative Matrix Factorization (NMF) has been found to learn and perform well in detecting incoherent topics [52]. As serviceable sub-types contain mostly posts on similar topics, it makes the topics more incoherent i.e. discuss diverse topical words associated with each topic. Next, we use Non-negative Matrix Factorization to identify optimal top topics (k) in serviceable tweets and iteratively evaluate k to minimize Frobenius norm. The resultant optimal value of k is found to be 5 for serviceable sub-types (See Table 2).

LDA topic	Vocabulary
Traffic congestion	Traffic, road, signal, bus, people, turn, vehicles, junction, stop, time, jam
Shared photos websites	com, www, facebook, https, videos, traffic, http, type, old, photos, job
Appreciation	signal, great, good, taking, act, ask facebook ,post ,news ,action
Question posed	asked, rules, Police, traffic, vehicle, number sir, said, car, know, time
Places	PS, telangana, state, hyderabad, city, nagar, afzalgunj, narayanguda, chaderghat, ousity
Fines issued	Challan [fine charged], violation, sir, bike, vehicle helmet, riding, traffic, like, documents
Cyber crime	Police, city, cyber, crime nampally, complaint, better, safe
NMF Topic	**Vocabulary**
Police incorrect decision	Police, asked, said, constable, taken, public wrong, driving, pay, vehicle, come, way
Awareness	documents, bike, know, mobile, rules need, people, let, share, helmet, circle, year
Dangerous driving complains	wrong, dangerous, action, driving, turn, going junction, coming, time, main, help, green
Fines issued	Vehicle, challan [fine charged], number, violation fine, documents, driving, guys, stopped, pay
Parking issues	Parking, people, bus, stop, parked, time, action notice, request, police, check, got, pollution, jams

Table 2: Extracted topics from annotated posts. LDA vocabulary includes content from serviceable and non-serviceable posts; NMF includes topics from serviceable sub-types.

SERVICEABLE VERSUS NON–SERVICEABLE REQUESTS

RQ1: Characterizing posts that get a police response

Per RQ 1(a) and 1(b), we begin by characterizing the differences in Serviceable and Non-Serviceable requests in terms of six content-based measures defined above. We further characterize the difference in posts based on the type of expected response by police (Serviceable sub-types).

Emotional Attributes: Table 3 presents the average value of residents' emotional states, intensity, and valence expressed in their posts. We find that serviceable requests show significantly higher value of negative emotional states i.e. "anger" (+15.38%), "disgust" (+47.8%), "fear" (+60%), and "sadness" (+10%) in comparison to non-serviceable requests. These posts show significantly higher arousal and valence than non-serviceable requests (See Table 3). Anxiety is also

[6] http://mpqa.cs.pitt.edu/opinionfinder/
[7] http://nlp.stanford.edu/software/tagger.shtml
[8] http://www.alchemyapi.com/products/alchemylanguage

measured to be (+100%) higher in serviceable posts in comparison to non-serviceable posts. Presumably, these emotional states are experienced due to distress caused because of encounters with law and order situation.

Emotional States							
	Serviceable			Non-Serviceable			
	Mn	Std	Md	Mn	Std	Md	Man
Anger	**0.15**	0.13	0.12	**0.13**	0.17	0.09	**-3.43****
Disgust	**0.34**	0.25	0.27	**0.23**	0.27	0.13	**-3.88****
Fear	**0.24**	0.21	0.18	**0.15**	0.18	0.07	**-6.09****
Sad	**0.11**	0.10	0.08	**0.10**	0.14	0.05	**-5.45****
Joy	**0.06**	0.11	0.02	**0.11**	0.20	0.01	-1.34
Valence							
Pos	3.78	5.88	2.08	2.71	5.67	0.00	**-3.92****
Neg	2.3	4.34	0.63	2.03	6.35	0.00	**-3.32****
Anxiety	**0.18**	0.77	0.00	**0.09**	0.52	0.00	-1.58
Intensity							
Arousal	4.90	1.73	5.42	2.64	2.77	0.00	**-6.53****

**p<0.01 *p<0.05
Pos=Positive affect, Neg=Negative affect

Table 3: Mn (Mean), standard deviation (Std.), and Median (Md) of Emotional attributes across serviceable posts. Statistical significance tested using Mann-Whitney U (Man).

On comparing different sub-types of serviceable requests (See Table 4), we find that "anger" is highest *Forward* (0.17) subtype posts followed by *Give Solution* (0.16) sub-type. Posts where police could reply with a "Thanks", express maximum "Joy" (0.16) and "Fear"(0.28) in comparison to other sub-types. We find statistically significant difference between "Thanks" and all other three sub-types of serviceability for "anger" (Dunn Test with correction, p<0.05). On analyzing posts in this subtype, we find that residents mention their encounters with law and order situations (experiencing fear) where police came to their rescue, marked by increased joy. For instance, in a post, resident shared that

> *I salute [withheld], [withheld] of Police (East), who did a great job of searching for the bodies of 3 missing children ([withheld]) by getting into sewage water at midnight 1:30AM at [withheld]*

"Negative valence" (4.62), "anxiety" (0.2) are maximum whereas positive valence is significantly less for *Needs more information* sub-type compared to others (Dunn Test with correction, p<0.05). Seemingly, these characteristics may indicate that residents when experiencing extreme emotions often do not mention complete information that is required for the police to help them. For example, complaining about a police officer, a resident shared feeling unsafe but did not mention the place or any identity of cops required to take action.

> *Two policemen tried to harass me today both of them were over speeding, and one of them abused me for taking a U-turn ......They were telling me that they will take a picture of my vehicle number plate and then show what the power of police.... #htp#imreallyfeelingunsafe*

Cognitive and Interpersonal Attributes: We observe that most of the *cognitive and Interpersonal attributes* can distinguish between serviceable and non-serviceable requests.

Emotional State						
		Frwd	Give	Thnk	Need	Krs
Anger	Mn	**0.17**	**0.16**	0.10	0.14	
	Std	0.14	0.14	0.10	0.12	13.66**
Joy	Mn	0.04	0.06	**0.16**	0.08	
	Std	0.08	0.08	0.19	0.11	44.42**
Fear	Mn	0.26	0.20	**0.28**	0.21	
	Std	0.22	0.17	0.24	0.18	3.97
Valence						
Pos	Mn	3.19	5.00	6.92	**2.21**	
	Std	4.41	8.22	8.55	3.40	24.90**
Neg	Mn	2.34	0.94	0.45	**4.62**	
	Std	2.64	2.06	1.32	8.19	71.86**
Anxiety	Mn	0.20	0.13	0.13	**0.20**	
	Std	0.80	0.59	0.85	0.76	2.37

**p<0.01 *p<0.05
Pos=Positive affect, Neg=Negative affect
Frwd = Forward; Give = Give Solution, Thnk = Thanks, Need = Need more Information

Table 4: Mean (Mn) and standard deviation (Std) for Emotional attributes across serviceable posts sub-types. Statistical significance tested using Kruskal Wallis (Krs). Values of other features discussed in Appendix Table 15.

On comparing *interpersonal focus* attributes (see Table 5), we find that serviceable requests show significantly higher use of 1st person singular pronouns (+9.1%) possibly indicating that serviceable requests are highly self-attention focused; also 3rd person pronouns (they (+18.2%)) are used significantly higher. First person singular pronoun shows a positive correlation with psychological distress [8]. In such posts, presumably, residents mostly express their own concerns.

Analysing serviceable subtypes, we observe that 1st person singular pronoun (2.56) is maximum in *Give Solution* subtype and minimum in *Thanks* (0.7) subtype. On analyzing posts, we find that *Give Solution* has maximum references to residents who need the police assistance to solve their own concerns. Contrastingly, *Thanks* includes post appreciating the police work which need not have references to one's own self but society as a whole. For instance,

> *Its really good by Hyderabad Traffic Police making traffic Awareness and teaching traffic rules to children.*

In Table 5, we observe that *cognitive process* measures can be used to decide between serviceable requests to non-serviceable requests. We find that "insight" (+46.2%), "cause" (+21.4%) and "certainty" (+64%) related words were significantly higher in serviceable posts than non-serviceable posts (See Table 5). Research shows that causal words are used while explaining reasons behind distressed experiences [54]. We conjecture that higher use of causal words explains the distress caused because of encountering a law and order situation.

Our observations on serviceable sub-types show statistically significant difference in all measures of *cognitive processes* (Kruskal-Wallis p<0.05, Dunn test with correction p<0.05). Thus suggesting, it can be informative to study these processes not only to differentiate between serviceable and non-serviceable posts but also to identify serviceable sub-type.

Linguistic Attributes: We find that serviceable posts showed higher (+40.2%) objective content than non-serviceable requests (See Table 6). Comparison of serviceable sub-types shows that objective statements are highest (3.47) in *Forward* posts and is least (1.9) in *Need more information* posts. Literature shows objectivity is strongly correlated with factual information [56]. We conjecture that requests high on objectivity contain more factual information on which the police can act upon.

Following this, we find that most dominant tense among serviceable posts is present tense (See Table 6). Presumably, our observation indicates that residents primarily request solutions for their ongoing concerns and issues. Analysis of serviceable sub-types shows that "present" tense is most frequently used in *Give solution* subtype. Further, past tense is most commonly used in *Need more information* followed by *Forward* subtype. We find that most of these posts, from *Need more information* and *Forward*, inquire about requests on which reply is awaited:

> *Cops driving the wrong side near [withheld] hotel .. what action will be taken against them?*

Further, we find that lexical density is significantly higher in serviceable posts than non-serviceable posts (see Table 6). The frequent occurrence of articles may indicate that residents make high attribution to things around them in their requests to police. Our observations show statistically significant difference between serviceable and non-serviceable posts for other measures of lexical density as well. However, analyzing the serviceable sub-types, we do not find any statistically significant differences between different sub-types for the use of lexical density terms. Thus indicating these to be useful measures for differentiating serviceable from non-serviceable posts.

		Serv	Non	Man	Frwd	Give	Thnk	Need	Krs
Interpersonal Focus									
i	Mn	**1.68**	**1.54**		1.61	**2.56**	**0.70**	1.80	
	Std	2.96	9.50	-3.80**	2.45	3.54	2.36	3.77	24.80**
she	Mn	**0.37**	**0.32**		0.30	0.25	0.37	0.62	
	Std	1.53	1.50	-1.06	0.97	1.04	1.44	2.74	2.63
they	Mn	**0.52**	**0.44**		0.63	0.32	0.26	0.53	
	Std	1.37	1.71	-2.19*	1.37	1.08	0.77	1.84	16.92**
Cognitive Mechanisms									
Cogn	Mn	11.68	8.63		11.68	14.4	11.24	9.70	
	Std	8.36	10.66	-4.37**	6.61	9.42	11.48	8.70	**16.26** **
Insght	Mn	**1.14**	**0.78**		1.01	1.9	0.75	1.13	
	Std	2.30	2.31	-3.54**	1.75	3.42	2.48	2.21	**16.96** **
Cause	Mn	**1.53**	**1.26**		1.55	1.63	1.42	1.50	
	Std	2.59	2.66	-2.44**	2.14	3.05	3.56	2.57	**9.93** *
Crtain	Mn	**1.46**	**0.89**		1.17	2.9	1.64	0.93	
	Std	4.19	3.55	-2.76**	1.97	7.28	3.92	5.03	**19.32** **

**p<0.01 *p<0.05
i = 1st person singular pronouns, she = 3rd person singular pronouns, they = 3rd person plural pronoun, Cong = Cognitive mechanisms;

Table 5: Interpersonal Focus & Cognitive orientation in serviceable posts and sub-types. These are statistically different using Mann-whitney U (Man) and Kruskal Wallis (Krs) test. Other features shown in Appendix Table 16.

		Serv	Non	Man	Frwd	Give	Thnk	Need	Krs
Objectivity									
Object -ive	Mn	**2.86**	**2.04**	-4.57**	**3.47**	2.64	2.07	**1.9**	37.70**
	Std	2.84	2.63		3.16	2.8	2.6	1.29	
Tense									
Past	Mn	1.75	0.81	-5.32**	1.88	1.68	0.78	2.14	18.49**
	Std	2.99	2.87		2.86	3.55	2.13	3.23	
Pres -ent	Mn	**7.34**	4.96	**-4.76** **	6.93	11.05	6.72	5.76	36.78**
	Std	6.29	7.14		4.47	8.03	9.02	5.48	
Lexical Density									
Article	Mn	4.99	2.47	**-6.20** **	5.49	4.1	4.18	4.91	3.45
	Std	4.64	4.53		4.25	4.27	5.31	5.31	
Verb	Mn	10.46	6.38	**-5.85** **	10.26	14.56	8.48	8.91	5.90
	Std	7.26	8.42		5.41	8.92	9.67	6.97	
Aux -verb	Mn	6.55	3.45	**-6.69** **	6.56	10.64	3.51	5.14	6.23
	Std	5.58	6		4.14	8.23	4.93	4.68	

**p<0.01 *p<0.05
Auxverb=auxiliary verb; Mn=Mean, Std=Standard Deviation
Frwd = Forward; Give = Give Solution, Thnk = Thanks, Need = Need more Information, Serv = serviceable, Non = Non-serviceable;

Table 6: Linguistic attributes in serviceable posts and its sub-types with statistical significance using Mann-whitney (Man) and Kruskal Wallis (Krs) test. Median and other features reported in Appendix, Table 17.

Question Asking Attributes: On analyzing question asking attribute as a binary feature, we find that 38.18% serviceable posts contain question-centric words in comparison to 33.62% non-serviceable posts (See Table 7). The difference in use of question-centric words is not statistically significant between serviceable and non-serviceable posts. While comparing sub-types, we find statistically significant difference between 4 sub-types regarding the use of question-centric words (χ^2, Phi = 0.384, p<0.05). For instance, almost 80% of posts in *Give Solution* use question-centric words followed by *Need more information* subtype with only 39.42% question-centric words. Qualitative analysis shows that *Give Solution* sub-type mostly consists of general queries regarding rules such as maximum fine that can be levied for any violation, which police can answer immediately on reading the post, e.g.,

> *Dear Admin, ...I had sold the 2-wheeler ..... when I went to echallans.org site and queried my vehicle there were a number of pending challans* [fine charged in case of violation] *which were clearly violated by the new owner........ How can I force the person to transfer the vehicle in his name?*

Entity-Based Attributes: Comparing number of *entities* mentioned in a post, we find that non-serviceable posts contain almost 23.17% higher the number of entities in serviceable posts (See Table 7). On qualitative analysis of non-serviceable posts, we find that residents mark their post to all possible police departments of a state thus increasing the number of entity count. We conjecture that residents show such behaviour as they want to propagate their content through these popular channels to get a response from some police organization. For example, below is a post where resident has tagged many police organizations to get attention:

> *WWW.HITECHNEWS.BIZ ...... Konaraopet PS PS Habeeb Nagar, Hyderabad City, Telangana State Kothapet PS PS Shahinayathgunj.........*

Table 7 shows that, among serviceable sub-types, *Give Solution* sub-type contain least number (0.86) of entities on an average per post followed by *Need more information* sub-type (1.14). As entity represent a person, place, or organization contained in a post, low entity count in *Need more information* sub-type may indicate that resident is not sharing complete details. For instance, in the post below the resident did not provide information about place and time making it difficult for police to act on it.

> *Dear Hyderabad Traffic Police, Vehicle parked at NO PARKING place... take necessary action.*

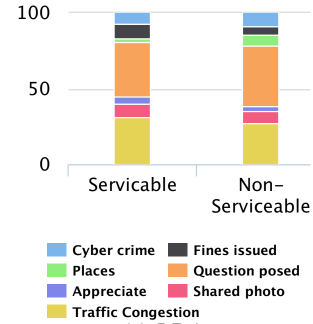
However, this may not hold true for *Give Solution* sub-type as most of the post in this sub-type are queries regarding rules or other facts containing limited references to different entities.

	Frwd	Give	Thnk	Need	χ^2	Serv	Non	χ^2
Question Asking Attributes								
%	28.32	**79.55**	25.00	**39.42**	80.93**	**38.18**	**33.63**	0.83
N	286	88	72	104		550	113	
Entity Count								
	Frwd	Give	Thnk	Need	Krs	Serv	Non	Man
Mn	1.89	**0.86**	1.72	**1.14**	37.22**	**1.56**	**2.04**	-3.20**
Std	1.86	1.20	2.00	1.34		1.75	4.51	

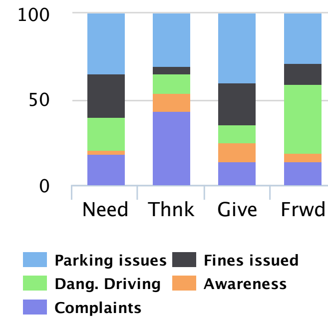
Table 7: Percentage of posts containing Question attributes and number of entities in serviceable posts and its sub-types with statistical significance **p<0.05. Mn=Mean and Std = Standard deviation

Topical Attributes: Our observations show statistically significant difference between serviceable and non-serviceable requests for topic labels associated with each post. Figure 1 (a) shows the most frequent topic in serviceable requests is "question posed to police" with frequent occurrence of words such as "asked", "rules", "said", and "know" (See Table 2). This topic is followed by information regarding "traffic congestion" caused due to nonfunctional signals, buses stopping anywhere on roads and complaints with autos drivers; contrastingly, the least frequent topic in non-serviceable posts refers to "appreciation" posts.

Further, on comparing serviceable sub-types, we find that "complaints regarding cops and decision with which residents don't agree" are most frequently mentioned in *Thanks* followed by *Forward* subtype. *Give solution* subtype shows the maximum occurrence of "parking issues" such as parking, wrong followed by "fines issued" sub-type (See Figure 1 (b)).



(a) LDA



(b) NMF

Figure 1: Number of topics using LDA and NMF for serviceable and its sub-types. Most frequent topic is question posed to police (Complaints represents complaints against cops incorrect decisions).

	Total N	N of Events	Censored N	% Censored
Forward	286	**182**	104	36.40%
Give Solution	88	53	35	39.80%
Thanks	72	5	67	**93.10%**
Need More Info.	104	60	44	42.30%
Serviceable	550	**300**	250	45.50%

Table 8: Number of posts that received responses (N of Events) and censored event showing posts that did not get response from the police.

We find statistically significant difference in the occurrence of four topics among all serviceable sub-types. Figure 1 demonstrates the distribution of different topics in serviceable (with its sub-types) and non-serviceable requests indicating it to be an effective measure for categorisation in our kind of task.

RQ2: How long does police take to respond?

In our annotated dataset, we find that among all the serviceable requests almost 55% (i.e. 300) got a response from police whereas none of the non-serviceable posts got any response. Comparing serviceable sub-types, we observe that 93.10% posts in *Thanks* sub-type did not receive a response from police. Posts in *Forward* sub-type received the maximum number of responses from police (63.6%, 182 posts). Table 8 summarizes the number of posts that did not receive police responses.

Next, we apply survival analysis to explore and estimate the police response time to serviceable posts and its subtypes. Survival analysis is a statistical method which examines the influence of time-related events such as estimating the probability that an event of interest occurs as a function of time [2, 40]. For our study, we define an event of interest as receiving a response from police. Survival analysis, in comparison to other statistical methods, offers an advantage that it considers truncated nature of events i.e. events which may not occur during the observation period of study but may occur after that. This technique is equipped to deal with such scenarios where the event of interest occurs at varying times. Thus taking into account such posts which do not receive a response from police during observation period i.e. data collection phase. We use Kaplan-Meir estimator for survival analysis, a nonparametric method, used to estimate the probability of survival past given time points.

For our purposes, we define the following terms:

- *Survival Time:* This is the time until the event of interest occurs i.e. time until a post receives a reply from the police. We define time intervals in minutes. Time when the police reply to a post is the *observed event*.
- *Censoring Event:* This consists of those posts which did not receive a reply during our observation period. However, these posts may receive a reply from police in future. For such posts, the exact survival time is not known and are called censored events. For a post that is not replied by police, survival time is the entire data collection phase from the time the post was created.
- *Survival Probability:* This gives the probability that a post remains unanswered by police longer than some specific time (t) given by survival function S(t). Further, to quantify the size of effect, we use *mean and median survival time* from our analysis. These values represent how long it takes for police to reply a post on an average.

The Kaplan-Meier estimator (See Table 9) shows that the median time to answer serviceable posts is 2,000 minutes (33 hours 20 min.). On comparing survival median for different sub-types of serviceable posts, we find after 1,226 minutes (20 hours 26 min.), 50% of the posts have received a reply in *Give Solution* sub-type where as it took 1,696 minutes (28 hours 16 min.) to achieve the same in *Needs more information* subtype (See Table 9).

	Mean Est.	Std. Err.	Median Est.	Std. Err.
Forward	63731.52	4933.47	1280	122.89
Give Solution	63862.87	8285.23	**1226**	507.03
Thanks	**161588.39**	5184.93	.	.
Need More Info.	68178.46	7649.03	**1696**	627.18
Serviceable	79616.12	3674.20	**2000**	.

Table 9: Means and Medians for Survival Time of different serviceable posts and sub-types. Survival median cannot not be calculated if censored data points are more than 50%.

Figure 2 shows survival curve as estimated (non-parametric) by Kaplan-Meier. Curve with lower probabilities represents shorter response time from the police which in our case is *Give Solution*. Our observations show statistically significant difference between all four sub-types of serviceable posts using Log Rank (Mantel-Cox) test (χ^2=57.03, df=3, p<0.005).

The maximum time for police to reply to a post is for *Thanks* sub-type where survival mean was maximum 161,588.39 minutes (2693 hours) (See Table 9). We find that censored posts percentage is 93.10% for *Thanks* sub-type showing that most of these posts did not get a reply during the observation period. These observations may indicate that the police considers it essential to reply to posts that can be *given solution* immediately followed by *Forward* sub-type and thus takes the

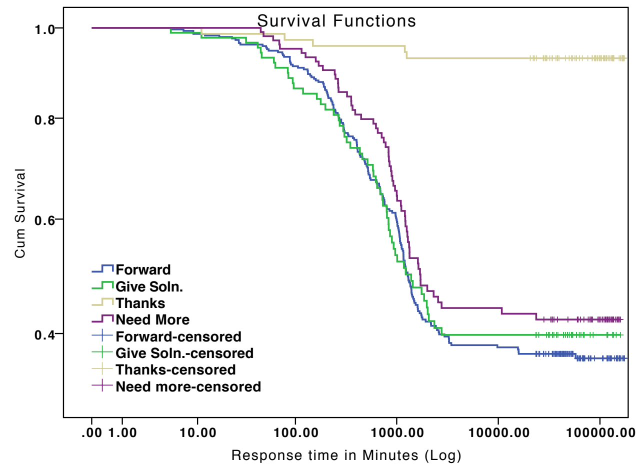


Figure 2: Survival curve as estimated (non-parametric) by Kaplan Meier for serviceable sub-types. Y-axis shows Cumulative Survival (Cum Survival).

least time to reply to such posts. On the other hand, *Thanks* consists of appreciation posts which can take some time to be replied to.

We find that survival median for posts from *Need more info.* sub-type is 38.33% greater than survival means of *Give Solution* and 32.50% greater than *Forward* sub-type. We conjecture that such delay may be caused because of police procedural steps involved in addressing the requests.

Next, we show that survival analysis can be used to update the expected response time in case a post does not receive a response from police in the estimated time. For instance, in *Forward* sub-type, if a post does not receive a reply in first 90 minutes, then the probability that the post will receive a reply in next 98 minutes would be (1-(.084 / .126))*100 = 33.34% (See Table 10). These values can be calculated for other subtypes to keep residents updated about when they are expected to receive a response.

Minutes	Survival Prob.	Cumulative Prob.	SE
090	0.916	**0.084**	0.016
121	0.906	0.094	0.017
133	0.895	0.105	0.018
188	0.874	**0.126**	0.020

Table 10: The cumulative probability that the post (Forward Sub-type) will be waiting for a police response.

RQ3: Modeling and Predicting Serviceable posts

As per RQ3(a) and RQ3(b), we now focus on building models which when given to a classifier can identify if a post is serviceable or not. Further, this model is fine-tuned to understand differences between serviceable sub-types. The model makes use of the content based measures – emotions, cognitive attributes, linguistic, question posed, entity and topical attributes discussed in RQ1. For RQ3(a), we formulate the identification task as a binary classification problem predicting whether a post is serviceable or not.

Algorithm	Precision	Recall	F1	Accuracy
RF	**0.89**	**0.97**	**0.85**	**0.87**
LG	0.90	0.82	**0.77**	**0.76**
ADT	0.88	0.96	0.80	0.86
DT	0.90	0.84	0.78	0.77
GBC	0.88	0.94	0.83	0.84

Table 11: Performance of different algorithms to correctly identify serviceable posts.

Final Values	R^2	Deviance	Precision	Recall	F1
Emotion	0.23	437.88	0.87	0.96	0.80
Linguistic	0.19	401.83	0.87	0.96	0.76
Cognitive	0.08	417.56	0.87	0.83	0.70
LDA,question,entity	0.03	129.21	0.87	0.70	0.70
Bag-of-words	0.53	260.07	0.91	0.84	0.79

Table 12: Goodness-of-fit for different algorithms to correctly identify serviceable posts.

To identify the best model for our task, we explore five different classification algorithms – Random Forest (RF), Logistic Regression (LR), Decision Trees (DT), Adaptive Boosted Decision Trees (ADT), and Gradient Boosting Classifier (GBC). Since, our classification task involves imbalanced dataset for different sub-types, we use balanced class weights. Research shows that balanced class weights perform better for classifying imbalanced classes as these automatically adjust weights inversely proportional to class frequencies in the input data [4]. Further, we use ten-fold cross-validation approach to evaluate our classification models. This method iteratively uses 90% data to train the classifier and rest 10% for testing the data. After ten iterations, the result shows the average accuracy and other performance measures. We also select five most important features using Chi-squared stats of non-negative features for classification tasks.

Table 11 shows the accuracy, precision, recall and F1 measure for our model in different algorithms. We find that Random Forest gives the best performance with F1 measure = 0.85 and accuracy of 87% while Logistic regression shows least F1 measure (value = 0.77) and accuracy of 76%. The model that includes all our suggested content based attributes performs best-showing increase of 0.06 (6%) in F-Measure than the baseline bag-of-words model. We report pseudo-R-squared values from logistic regression for our different measures (content based attributes) as a goodness-of-fit statistic. These values are supported with F1, precision, and recall from Random Forest that performs best for our model. As seen from pseudo R-squared, content attributes such as emotions and linguistic attributes are highly predictive of serviceable posts in addition to bag-of-words model (See Table 12).

Classifying serviceable sub-types

All serviceable sub-types consist of some information which police may need to act upon, however, there are subtle differences among the four sub-types. The content characteristics based model that separates serviceable posts from non-serviceable ones (described above) when used to classify the serviceable sub-types gives an accuracy of approximately 58% only. Therefore, per RQ3(b), we focus on improved classification of serviceable sub-types and present in-depth analysis of different models that we fit using Step-Entry logistic regression to understand the relative importance of these features. This analysis is based on stepwise multinomial logistic regression; specifically we use *forward entry* method. Independent variables in this regression model are the six content-based measures and dependent variable are the four sub-types – *Forward*, *Give solution*, *Thanks* and *Need more information*. We begin our analysis with a null model and add new measures step-by-step. We further rely upon different performance criteria such as deviance / log likelihood to evaluate these models. [9] This process helps to determine the importance of each measure in predicting the sub-types.

Table 13 shows the increase in performance of different measures. We fit a bag-of-words model as the null model to start with the evaluation process. This model uses frequency of words as a feature for predicting sub-type labels and shows a deviance of value 1,305.72. The corresponding R-square value indicates that this model explains about 2.3% variance in the data (See R-squared value in Table 13). Next, we evaluate the influence of emotional attributes in identifying different serviceable sub-types. Our observations show that adding emotional attributes to the null model slightly improves the performance by decreasing deviance (See Table 13). As shown in Table 13, this Model 1 explains about 15.6% of the variance in data in comparison to 2.3% that is described by the null model. Further, Model 1 accounts for reducing deviance significantly to 1,127.58 (178.14 less). Some of the distinguishing emotional attributes that act as better predictors of the serviceability sub-types are emotional states including sadness, fear, and joy. This is in agreement with our observations discussed in RQ1 showing that there exist significant differences in emotional state and emotional valence when comparing serviceable sub-types.

We included cognitive and interpersonal attributes to Model 1 to derive Model 2 and evaluated the performance of same. This model further reduces the deviance to 1,068.82 improving the performance. As seen from R-square values, these features explain about 20% variance in the data (See Table 13). Despite this improvement, we see that only a few linguistic features such as tentativeness and pronouns act as the major contributors among various cognitive attributes. Logistic Regression analysis shows that 1st person singular pronouns have statistically significant intercept values for both *give solution* and *need more information* category, whereas these values are not significant for *thanks* sub-type. Thus showing the importance of 1st person singular pronouns in classifying the different sub-types.

Through Model 3, we explore the performance of linguistic features – subjectivity, lexical density, tense use (past, future and present), and part-of-speech tags in predicting the sub-types labels when included to existing model. We find that

[9]We use deviance as the goodness of fit for a model as this measure does not have any direct predictor parameters such as R-squared in OLS regression. We suggest that these values should, therefore, be interpreted with caution.

	Effect(s)	Give soln.			Thanks			Need more						
		B	SE	Sig.	B	SE	Sig.	B	SE	Sig.	Dev	Psuedo R^2	Likelihood	AIC
	Intercept	0.158	0.402	0.695	-0.542	0.457	0.236	-0.883	0.411	0.032				
	Fear	-1.761	0.586	0.003	0.364	0.566	0.521	-1.114	0.545	0.041				
+Model 1	Joy	1.553	1.362	0.254	7.386	1.245	0.000	4.979	1.220	0.000	**1127.59**	0.156	208.21	1181.59
	Sad	2.134	0.987	0.031	-6.049	1.819	0.001	2.475	0.958	0.010				
	Negative	-0.257	0.059	0.000	-0.403	0.101	0.000	0.084	0.027	0.002				
	Intercept	-0.842	0.388	0.030	-0.408	0.401	0.308	-1.314	0.384	0.001				
+Model 2	Pronoun	0.091	0.017	**0.000**	-0.007	0.021	0.753	0.053	0.017	**0.002**	**1068.82**	0.200	266.98	1134.82
	Tentative	0.157	0.037	0.000	-0.242	0.081	0.003	0.018	0.043	0.682				
	Inclusion	-0.085	0.038	0.024	0.062	0.035	0.076	-0.112	0.036	0.002				
	Intercept	-1.324	0.416	0.001	-0.108	0.403	0.788	-0.430	0.361	0.233				
	Negate	-0.265	0.065	0.000	-0.120	0.081	0.138	-0.019	0.046	0.680				
+Model 3	Auxvrb	0.093	0.02	0.000	-0.121	0.031	0.000	-0.051	0.024	0.031	**984.92**	0.263	350.88	1080.92
	P-O-S	1.752	0.753	0.020	0.612	0.810	0.450	-1.583	0.664	0.017				
	subjective	-0.797	0.181	0.000	-0.314	0.165	0.057	-0.410	0.138	0.003				
	Intercept	-0.668	0.535	0.211	-0.258	0.547	0.638	-1.236	0.506	0.015				
	Question	-2.880	0.284	0.000	-0.861	0.293	0.003	-1.540	0.24	0.000				
+Model 4	entity	-0.419	0.099	0.000	0.088	0.073	0.228	-0.102	0.074	0.167	**902.79**	0.324	433.01	1004.79
	Datetime	1.452	0.449	0.001	1.099	0.421	0.009	2.256	0.403	0.000				
	Jobtitle	5.237	1.439	0.000	4.496	1.626	0.006	3.411	1.511	0.024				

Table 13: Performances of forward-entry logistic regression on different models. The effects (variables) are added incrementally and performance is reported for each model. The reference category is *Forward*.

amount of subjectivity, auxiliary verbs, negations and different Part-Of-Speech used in a post are among the prominent features to distinguish among subtypes of serviceable posts. This model reduces the deviance to a value of 984.92 and explains about 26.3% of the variance in the data (See Table 13). However, we find that many attributes such as tenses (present and future) and lexical terms (verbs and adverbs) did not help much in differentiating the four sub-types of serviceability.

The last model (Model 4) examines and incrementally adds four attributes – presence of questions, date & time, entities and topic associated (using Non-negative Matrix Factorization) to the existing model. This model shows significant improvement in the performance of the classifier by reducing the deviance to 902.79 i.e. 82.13 points less than the previous model (See Table 13). We find that attributes such as the presence of a question, date, time, and entity count (number of entities in a post) are reliable predictors of serviceable sub-types. This model explains 32.4% of the variance in the model. However, we find that topic associated with a post does not appear in the top deterministic features. This could be because some of these topics appear consistently in multiple subtypes. Thus reducing the significance of Topic as a feature in comparison to other features.

Finally, based on the feature exploration using multinomial logistic regression discussed above, we gain fair insights about how different attributes contribute to deciding labels of sub-types. Next, we use this knowledge to understand how well classifiers can predict the sub-types of serviceable posts (Table 14). Similar to the methodology used in RQ3(a), we use four algorithms for our experiments including – Random Forest (RF), Logistic Regression (LR), Adaptive Boosted Decision Trees (ADT), and Gradient Boosting Classifier (GBC). Table 14 summarises the performance (obtained using 10 fold cross-validation) of these different algorithms on our models

	Accuracy	F1	Precision	Recall
RF	0.61	0.57	0.58	0.61
LR	**0.65**	**0.61**	**0.63**	**0.65**
ADT	0.61	0.57	0.63	0.59
GBC	0.61	0.57	0.60	0.61

Table 14: Performance of classification for sub-types of serviceable posts as measured by mean values of Accuracy, F1, Precision, and Recall over 10-fold Cross Validation.

(See Table 14). We find that Logistic Regression gives the best performance with F1 measure = 0.61 and accuracy of 65%. Using our observations, we are able to improve the accuracy by 7% from the initial 58% that is achieved using the same features as for serviceable versus non-serviceable posts.

DISCUSSION

Our work demonstrates the viability of using content generated on Facebook police pages as an instrument to quantify and characterize resident's requests that should elicit police response (called *serviceable requests* in our context). We describe the subtleties of serviceable requests and response provisions available in the context of policing through social media. This section reflects on our results and discusses contributions of our work contributes to prior knowledge on police use of social media during crises, data-driven policing, and policing practices in urban communities.

This work contrasts with prior work exploring the use of social media by police in that we study social media use for day-to-day policing needs in comparison to the well-explored scenario of crisis. Highlighting social media's role, research shows that residents can become a valuable resource for gathering crisis information [51, 57]. Prior work also suggests that a non-responsive social media presence can damage the residents' willingness to engage with and trust in the police [47, 48]. Also, most of the existing studies explore how police push information to or extract relevant information from residents to address safety critical situations [37]; whereas our work looks at what residents tend to share with police in day-to-day life. Our study shows that police respond more to serviceable posts than they do to general communication. We believe, social platforms facilitating serviceable requests should allow for design choices that help police monitor serviceable requests and manage a timely response.

Social media use for policing results in a) large volume of content sharing by residents and b) greater expectations among residents for how police may respond and manage the shared content [25, 48]. To address the resident's expectations, research suggests aiding traditional policing methods with automated / data-driven interventions and making social media streams 'listenable' to residents' concerns [25, 48]. In this paper, we develop an automated approach for detecting posts that should elicit police response that may help make social media streams more listenable for resident's concerns. As a significant contribution, we present an expert (police officers) annotated dataset useful in assessing serviceable post. The six broad measures proposed in our work help identify the textual constructs that can characterize serviceable requests and act as predictors to recognize incidences of interest for police. Through our work, we hope to help the police departments to use social media effectively for reducing the time taken to address residents' concern. In countries such as India, for every 100,000 residents there are only 130 police officers deputed. In lack of adequate human resource, it becomes imperative to resort to automated data-driven policing approaches that can aid police departments to provide timely response to the citizens.

Community policing theory suggests to involve community members in the problem-solving process to prevent crime and create safer societies [33, 47]. Using data-driven techniques, our work makes an attempt to characterize and quantify different resident's concerns expressed on social media that can complement police documentation of the various incidences reported in the communities. We show that residents top concerns include complaints regarding police incorrect decision, fine charged, traffic violations, and dangerous driving practices. Posts include factual information such as date, time, and location of event helping police identify areas of concerns. Taking cognizance of prominent constituents' concern and unsafe regions can help police plan their resources better to provide improved safety. Future research may explore ways to highlight important topics emphasized by the community to ensure that they receive proper attention during police organization's resource planning.

In addition to mentioning factual information, we measure resident's reactions in a fine-grained manner. Prior work shows that police should account for citizen reactions experiencing a law and order situation [15]. We find requests that police respond to are marked by high negative emotions and are highly self-focused. Using our proposed measures, these reactions can be observed on a large scale, through textual posts from Facebook; our work can be extended to other social media too, keeping in mind the basic differences in the different social media services. We believe that various aspects of shared data (e.g., emotions and interpersonal attributes) improve the understanding from mere factual information (incidences reported on conventional helplines) to a more nuanced understanding of psychological aspects such as emotions and social changes. This understanding may have implication on designing early warning systems that indicate targets where improvement in emotional efficacy and social cohesion can be encouraged to enhance emotional support to residents experiencing safety issues.

Previous research shows that by posting concerns to a public forum like Facebook, residents hold the police publicly accountable [25, 33, 48]. We explored different types of response that police may give to residents. In our observations, we find that the maximum number of posts from 'forward' (63%) received police response followed by 'need more information'. Police are fastest in replying to 'Give Solution' followed by 'forward'. 'Need more information' that did not have all information took longest of the three to receive a reply (Forward, give solution and need more information). These responses may indicate that police makes an attempt to address resident's accountability concerns on the part of the police and makes efforts to foster trust by addressing their queries.

Online environments lack other signals of communication (for instance, physical proximity and eye gaze) that may ascertain police presence thus comments or reply are an important feedback to increase feelings of connectedness in online communities [19]. Criminology theory suggests that police presence and visibility in the community can have an impact on resident's behavior and fear of crime [1]. Our work shows that textual characteristics can help gauge if a resident's post should receive a police reply and probable duration in which response would come. We believe a feedback system could be integrated into the design of social media platforms that informs a resident about the likely time duration that police would take to respond to service requests, thus augmenting the current feedback system. This feedback may also encourage residents to use social media frequently for transmitting information about the neighborhood and give residents a sense of personal contribution in the community.

Limitations and Future Work

Although we see potential use of social media to develop technologies for improved landscaping of safety in urban cities, we understand that these technologies may not work as standalone solutions. Rather these solutions may complement existing methods and become part of broader detection systems about resident's psychological and social responses. We also suggest caution in interpreting findings of our work. Our work shows that different service requests get a varied police response. However, we do not make any claims about whether these responses meet the expectation and actual need of the residents. We recommend this as an important step for future studies which could use interview and survey methods to evaluate the usefulness of these responses.

In our study, we primarily perform textual analysis; however, social media streams consist of other data like images and videos that could be useful to characterize serviceable posts. Future work can explore visual content as binary features and use computer vision techniques such as objects detection and text recognition in the visual content to identify serviceability.

We believe that our suggested measures can help account for concerns of residents who willingly express their views on social media. However, we understand that the willingness to share data may differ among communities based on their broader cultures. For instance, communities that may have a strong culture of community policing may be more willing to engage in collective action through these methods. However, communities where historically trust has been a challenge between police and residents may need to be persuaded more to participate. Future work could explore the influence of culture variations in this context. Further, to complement our study, we recommend future work to compare our findings for rural communities where Facebook is showing fast penetration. Rural communities may show varied concerns than discussed in urban areas. Thus more exploration is required to understand how our suggested measures calibrate in such contexts. It may also be worthwhile to examine if factors such as demography and education influence police responses.

CONCLUSION

In this paper, we explored what posts should elicit police response on Facebook pages of Indian police departments. We study six different textual attributes that can help identify such serviceable posts. Our observations show that serviceable posts were marked by high negative emotions, factual content and references to objective content. Further, police response varies based on different type of serviceable posts and almost 50% of serviceable posts receive response within 33.34 hours. We find that textual features identified in our work can help predict if a post should elicit police response. To this end, we explore different statistical and prediction models. Our findings bear implications in the design of data-driven technologies to provide timely help and support to citizens for improved safety and well being.

ACKNOWLEDGEMENT

We would like to thank TCS research for funding the project. Also, we would like to thank the members of Cybersecurity Education and Research Centre (CERC) and Precog who gave us continued support throughout the project; special thanks to Siddhartha Asthana, Megha Arora and Indira Sen.

REFERENCES

1. James F Albrecht, Martha Christine Dow, Darryl Plecas, and Dilip K Das. 2014. *Policing Major Events: Perspectives from Around the World*. CRC Press.
2. Siddhartha Asthana, Pushpendra Singh, and Parul Gupta. 2015. Survival Analysis: Objective assessment of Wait Time in HCI. In *In Proc. CHI 15*.
3. Valerio Basile, Elena Cabrio, Serena Villata, Claude Frasson, and Fabien Gandon. 2016. A Pragma-Semantic Analysis of the Emotion/Sentiment Relation in Debates. In *4th International Workshop on Artificial Intelligence and Cognition*.
4. Chao Chen, Andy Liaw, and Leo Breiman. 2004. Using random forest to learn imbalanced data. *University of California, Berkeley* (2004).
5. Munmun De Choudhury, Andres Monroy-Hernandez, and Gloria Mark. 2014. “Narco” Emotions: Affect and Desensitization in Social Media during the Mexican Drug War. In *Proc. CHI ’14*. ACM.
6. Camille Cobb, Ted McCarthy, Annuska Perkins, Ankitha Bharadwaj, Jared Comis, Brian Do, and Kate Starbird. 2014. Designing for the Deluge: Understanding & Supporting the Disturbed, Collaborative Work of Crisis Volunteers. In *Proc. CSCW ’14*.
7. Adam Crowe. 2010. The elephant in the JIC: The fundamental flaw of emergency public information within the NIMS framework. *Journal of Homeland Security and Emergency Management* 7, 1 (2010).
8. Wendy D’Andrea, Pearl H Chiu, Brooks R Casas, and Patricia Deldin. 2012. Linguistic predictors of post-traumatic stress disorder symptoms following 11 September 2001. *Applied Cognitive Psychology* 26, 2 (2012), 316–323.
9. Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting Postpartum Changes in Emotion and Behavior via Social Media. In *Proc. CHI ’13*.
10. Munmun De Choudhury, Scott Counts, Eric J Horvitz, and Aaron Hoff. 2014. Characterizing and predicting postpartum depression from shared facebook data. In *Proc. CSCW ’14*. ACM, 626–638.
11. Louis M. Dekmar. 2015. Handling Citizen Complaints through Proactive Methodology. http://www.policechiefmagazine.org/magazine/index.cfm?fuseaction=display_arch&article_id=2052&issue_id=42010. (Sept. 2015).
12. Sebastian Denef, Petra S. Bayerl, and Nico Kaptein. 2013. Social Media and the Police - Tweeting Practices of British Police Forces during the August 2011 Riots. (2013).
13. Department of Economic and Social Affairs, UN. 2015. World Population Prospects - The 2015 Revision. https://esa.un.org/unpd/wpp/publications/files/key_findings_wpp_2015.pdf. (2015).
14. Department of Electronics and Information Technology. 2012. Framework & Guidelines for Use of Social Media for Government Organisations. https://www.negp.gov.in/pdfs/Social%20Media%20Framework%20and%20Guidelines.pdf. (2012).
15. John Drury and Steve Reicher. 2000. Collective action and psychological change: The emergence of new social identities. *British Journal of Social Psychology* (2000).

16. Express News Service. 2013. Is community policing need of the hour? http://www.newindianexpress.com/states/karnataka/article1430481.ece. (2013).

17. Zann Gill. 2012. User-driven collaborative intelligence: social networks as crowdsourcing ecosystems. In *CHI'12 Extended Abstracts*.

18. Andrew Goldsmith. 2015. Disgracebook policing: social media and the rise of police indiscretion. *Policing and society* 25, 3 (2015), 249–267.

19. Rebecca Gray, Nicole B Ellison, Jessica Vitak, and Cliff Lampe. 2013. Who wants to know?: question-asking and answering practices among facebook users. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1213–1224.

20. Aditi Gupta and Ponnurangam Kumaraguru. 2012. Credibility ranking of tweets during high impact events. In *Proc. PSOSM colocated with WWW '12*. ACM, 2.

21. Thomas Heverin and Lisl Zach. 2010. Twitter for City Police Department Information Sharing. In *Proc ASIST '10*.

22. Lyn Hinds and Kristina Murphy. 2007. Public satisfaction with police: Using procedural justice to improve police legitimacy. *Australian & New Zealand Journal of Criminology* 40, 1 (2007), 27–42.

23. Amanda L Hughes and Leysia Palen. 2012. The evolving role of the public information officer: An examination of social media in emergency management. *Journal of Homeland Security and Emergency Management* 9, 1 (2012).

24. Amanda L Hughes, Leysia Palen, Jeannette Sutton, Sophia B Liu, and Sarah Vieweg. 2008. "Site- Seeing" in disaster: An examination of on-line social convergence.. In *Proc. ISCRAM '08*.

25. Amanda L Hughes, Lise AA St Denis, Leysia Palen, and Kenneth M Anderson. 2014. Online public communications by police & fire services during the 2012 Hurricane Sandy. In *Proc. CHI '14*.

26. Indian Police Department. 2015. Citizen Charter. http://www.bcp.gov.in/doc/BCP%20Citizen%20Charter.pdf, http://www.delhipolice.nic.in/citizen%20chart.pdf. (2015).

27. Dianne Jeans. 1993. The Relationship Between Police and Other Government Agencies: Recent Changes in Perspective in Queensland.

28. Andrea Kavanaugh, Edward Fox, Steven Sheetz, Seungwon Yang, Lin Tzy Li, Travis Whalen, Donald Shoemaker, Paul Natsev, and Lexing Xie. 2011. Social Media Use by Government: From the Routine to the Critical. (2011).

29. Cliff Lampe, Rebecca Gray, Andrew T Fiore, and Nicole Ellison. 2014. Help is on the way: Patterns of responses to resource requests on Facebook. In *Proc. CSCW '14*.

30. Mark Latonero and Irina Shklovski. 2011. Emergency management, Twitter, and social media evangelism. *International Journal of Information Systems for Crisis Response and Management* 3, 4 (2011), 67–86.

31. Murray Lee and Alyce McGovern. 2013. *Policing and Media: Public Relations, Simulations and Communications*. Routledge.

32. Dan A Lewis and Greta Salem. 1981. Community crime prevention: An analysis of a developing strategy. *Crime & Delinquency* 27, 3 (1981), 405–421.

33. Sheena Lewis and Dan A. Lewis. 2012. Examining Technology That Supports Community Policing. In *Proc. CHI '12*.

34. Claudia A. López and Brian S. Butler. 2013. Consequences of Content Diversity for Online Public Spaces for Local Communities. In *Proc. CSCW '13*.

35. Jalal Mahmud. 2016. Another Step Closer to Building Empathetic Systems. http://blog.alchemyapi.com/a-step-closer-to-building-empathetic-systems. (2016).

36. Stephen D Mastrofski, Jeffrey B Snipes, Roger B Parks, and Christopher D Maxwell. 2000. The helping hand of the law: police control of citizens on request*. *Criminology* 38, 2 (2000), 307–342.

37. Albert Meijer and Marcel Thaens. 2013. Social media strategies: Understanding the differences between North American police departments. *Government Information Quarterly* 30, 4 (2013), 343–350.

38. Mendoza, Marcelo and Poblete, Barbara and Castillo, Carlos. 2010. Twitter under crisis: Can we trust what we RT?. In *Proc. Social Media Analytics SOMA '10*.

39. Ines Mergel. 2014. Social Media Adoption: Toward a Representative, Responsive or Interactive Government?. In *Proc dg.o '14*. ACM, 163–170.

40. Rupert G Miller Jr. 2011. *Survival analysis*. Vol. 66. John Wiley & Sons.

41. Vinaya Naidu. 2014. Indian Police Leveraging Social Media Platforms And Messaging Apps To Maintain Law And Order. lighthouseinsights.in/indian-police-using-social-media.html/. (March 2014).

42. Leysia Palen and Sophia B Liu. 2007. Citizen communications in crisis: anticipating a future of ICT-supported public participation. In *Proc. CHI '07*.

43. Leysia Palen and Sarah Vieweg. 2008. The Emergence of Online Widescale Interaction: Assiatance, Alliance and Retreat. In *Proc. CSCW '08*.

44. Yan Qu, Philip Fei Wu, and Xiaoqing Wang. 2009. Online Community Response to Major Disaster: A Case Study of Tianya Forum in the 2008 China Earthquake. *In Proc 42nd Hawaii Int'l Conf. on System Sciences* (2009).

45. Daniele Quercia, Harry Askham, and Jon Crowcroft. 2012. TweetLDA: supervised topic classification and link prediction in Twitter. In *Proceedings of the 4th Annual ACM Web Science Conference*. ACM, 247–250.

46. Niharika Sachdeva, , Ponnurangam Kumaraguru, and Munmun De Choudhury. 2015. Social Media for Safety: Characterizing Online Interactions between Citizens and Police. *In Proc. HCI* (2015).
47. Niharika Sachdeva and Ponnurangam Kumaraguru. 2015a. Online Social Networks and Police in India - Understanding the Perceptions, Behavior, Challenges. *In Proc. ECSCW 15* (2015).
48. Niharika Sachdeva and Ponnurangam Kumaraguru. 2015b. Social networks for police and residents in India: exploring online communication for crime prevention. *In Proc. Dg.o 15* (2015).
49. Bryan Semaan and Gloria Mark. 2012. 'Facebooking' Towards Crisis Recovery and Beyond: Disruption as an Opportunity. *In Proc. CSCW* (2012), 27–36.
50. Irina Shklovski, Leysia Palen, and Jeannette Sutton. 2008. Finding Community through information and communication technology in disaster response. In *Proc. CSCW '08*.
51. Kate Starbird and Leysia Palen. 2011. "Voluntweeters": Self- organizing by digital volunteers in times of crisis. In *Proc. CHI '11*. 1071–1080.
52. Keith Stevens, Philip Kegelmeyer, David Andrzejewski, and David Buttler. 2012. Exploring topic coherence over many models and many topics. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.
53. Shobhit Sujay. 2015. 24x7 duty for police personnel in India: What forces across the world follow. http://www.news18.com/news/india/24x7-duty-for-police-personnel-in-india-what-forces-across-world-follow-1017408.html. (2015).
54. Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.
55. Tucson Police. 2016. Calls for Service. https://www.tucsonaz.gov/police/terms. (May 2016).
56. Sudha Verma, Sarah Vieweg, William J Corvey, Leysia Palen, James H Martin, Martha Palmer, Aaron Schram, and Kenneth Mark Anderson. 2011. Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency. Citeseer.
57. Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. 2010. Micro-blogging during two natural hazards events: What Twitter May Contribute to Situational Awareness. In *Proc CHI '10*. 1079–1088.
58. Amy Voida, Ellie Harmon, and Ban Al-Ani. 2012. Bridging Between Organisation and the Public: Volunteer Coordinators' Uneasy Relationship with Social Computing. In *Proc. CHI '12 (ACM)*.
59. Yi-Chia Wang, Robert Kraut, and John M. Levine. 2012. To Stay or Leave?: The Relationship of Emotional and Informational Support to Commitment in Online Health Support Groups. In *Proc. CSCW '12*.
60. Eric W Welch and Shelley Fulla. 2005. Virtual interactivity between government and citizens: The Chicago Police Department's citizen ICAM application demonstration case. *Political communication* (2005).
61. Jian Zhao, Liang Gou, Fei Wang, and Michelle Zhou. 2014. PEARL: An interactive visual analytic tool for understanding personal emotion style derived from social media. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference*.

APPENDIX

Additional Tables

		Emotional State					Valence			Intensity
		Anger	Joy	Fear	Sadness	Disgust	Pos	Neg	Anxiety	Arousal
Forward	Mean	0.17	0.04	0.26	0.10	0.34	3.19	2.34	0.20	5.15
	Std	0.14	0.08	0.22	0.10	0.25	4.41	2.64	0.80	1.23
	Med	0.12	0.08	0.22	0.10	0.25	2.17	1.90	0.00	5.40
Give Solution	Mean	0.16	0.06	0.20	0.14	0.37	5.00	0.94	0.13	4.62
	Std	0.14	0.08	0.17	0.11	0.23	8.22	2.06	0.59	2.02
	Med	0.14	0.03	0.15	0.11	0.32	1.54	0.00	0.00	5.42
Thanks	Mean	0.1	0.16	0.28	0.09	0.27	6.92	0.45	0.13	4.57
	Std	0.10	0.19	0.24	0.07	0.24	8.55	1.32	0.85	2.46
	Med	0.08	0.09	0.18	0.07	0.18	4.33	0.00	0.00	5.65
Need more Info.	Mean	0.14	0.08	0.21	0.14	0.36	2.21	4.62	0.20	4.70
	Std	0.12	0.11	0.18	0.12	0.26	3.40	8.19	0.76	1.94
	Med	0.12	0.03	0.17	0.10	0.31	0.00	1.58	0.00	5.34
Kruskal Wallis		13.66**	44.42**	3.97	23.58*	10.42	24.90**	71.86**	2.37	3.92
Serviceable	Mean	0.15	0.06	0.24	0.11	0.34	3.78	2.3	0.18	4.90
	Std	0.13	0.11	0.21	0.10	0.25	5.88	4.34	0.77	1.73
	Med	0.12	0.02	0.18	0.08	0.27	2.08	0.63	0.00	5.42
Non-Serviceable	Mean	0.13	0.11	0.15	0.10	0.23	2.71	2.03	0.09	2.64
	Std	0.17	0.20	0.18	0.14	0.27	5.67	6.35	0.52	2.77
	Med	0.09	0.01	0.07	0.05	0.13	0.00	0.00	0.00	0.00
Mann Whitney U		-3.43**	-1.34	-6.09**	-5.45**	-3.88**	-3.92**	-3.32**	-1.58	-6.53**

*p<0.01 *p<0.05
Pos=Positive affect, Neg=Negative affect

Table 15: Mean, standard deviation (Std.), and Median(Med) of Emotional attributes across serviceable posts and its sub-types. Results of differences and statistical significance tests using Mann-Whitney U and Kruskal Wallis Test.

		Cognitive				Interpersonal					
		Cgmch	Insight	Cause	Certain	i	we	you	she	they	ipron
Forward	Mean	11.68	1.01	1.55	1.17	1.61	0.45	0.79	0.30	0.63	3.77
	Std	6.61	1.75	2.14	1.97	2.45	1.11	1.55	0.97	1.37	3.04
	Med	11.81	0.00	0.72	0.00	0.00	0.00	0.00	0.00	0.00	3.46
Give Solution	Mean	14.4	1.90	1.63	2.9 0	2.56	0.82	1.11	0.25	0.32	6.86
	Std	9.42	3.42	3.05	7.28	3.54	1.87	2.98	1.04	1.08	8.77
	Med	12.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.55
Thanks	Mean	11.24	0.75	1.42	1.64	0.70	1.21	1.19	0.37	0.26	3.14
	Std	11.48	2.48	3.56	3.92	2.36	3.53	3.13	1.44	0.77	4.88
	Med	10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Need More Info.	Mean	9.7	1.13	1.5	0.93	1.80	0.58	0.88	0.62	0.53	3.44
	Std	8.70	2.21	2.57	5.03	3.77	2.08	1.98	2.74	1.84	4.22
	Med	10.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.47
Kruskal Wallis		16.26**	16.96**	9.93*	19.32**	24.80**	2.87	3.61	2.63	16.92**	17.64**
Serviceable	Mean	11.68	1.14	1.53	1.46	1.68	0.63	0.91	0.37	0.52	4.12
	Std	8.36	2.30	2.59	4.19	2.96	1.92	2.17	1.53	1.37	4.99
	Med	11.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.14
Non-Serviceable	Mean	8.63	0.78	1.26	0.89	1.54	0.73	1.29	0.32	0.44	2.39
	Std	10.66	2.31	2.66	3.55	9.50	2.85	3.81	1.50	1.71	4.29
	Med	4.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mann Whitney		-4.37**	-3.54**	-2.44**	-2.76**	-3.80**	-0.50	-1.68	-1.06	-2.19*	-5.20**

**p<0.01 *p<0.05
Cgmch = Cognitive mechanisms, i = 1st person singular pronouns, we = 1st person plural pronouns, you = 2nd person she = 3rd person singular pronouns, they = 3rd person plural pronoun, ipron = impersonal pronouns

Table 16: Cognitive orientation measures in serviceable posts and sub-types. Cognitive and Interpersonal focus are statistically different in serviceable posts and non-serviceable posts using Mann-whitney U Test.

		Objectivity		Tense			Grammer				
		subj	obj	past	prsnt	futr	artcl	vrb	axvrb	quant	numb.
Forward	Mean	0.77	3.47	1.88	6.93	0.68	5.49	10.26	6.56	2.13	0.54
	Std	1.23	3.16	2.86	4.47	1.40	4.25	5.41	4.14	2.71	1.13
	Med	0.00	3.00	0.26	6.79	0.00	5.21	10.34	6.63	1.41	0.00
Give Solution	Mean	0.24	2.64	1.68	11.05	0.75	4.10	14.56	10.64	2.47	0.33
	Std	0.53	2.8	3.55	8.03	1.64	4.27	8.92	8.23	4.08	1.19
	Med	0.00	2.00	0.00	9.6	0.00	4.09	14.29	9.15	0.00	0.00
Thanks	Mean	0.31	2.07	0.78	6.72	0.62	4.18	8.48	3.51	2.09	0.57
	Std	0.78	2.60	2.13	9.02	1.86	5.31	9.67	4.93	4.16	1.73
	Med	0.00	1.00	0.00	4.55	0.00	2.49	6.31	0.00	0.00	0.00
Need More Info.	Mean	0.38	1.90	2.14	5.76	0.61	4.91	8.91	5.14	1.37	0.53
	Std	1.00	1.29	3.23	5.48	1.82	5.31	6.97	4.68	2.45	1.91
	Med	0.00	1.00	0.00	5.41	0.00	4.17	9.38	5.26	0.00	0.00
Kruskal Wallis		37.18**	37.70**	18.49**	36.78**	10.53*	3.45	5.90	6.23	2.92	6.86
Serviceable	Mean	0.55	2.86	1.75	7.34	0.67	4.99	10.46	6.55	2.03	0.51
	Std	1.07	2.84	2.99	6.29	1.59	4.64	7.26	5.58	3.15	1.40
	Med	0.00	2.00	0.00	6.82	0.00	4.76	10.53	6.40	0.00	0.00
Non-Serviceable	Mean	0.19	2.04	0.81	4.96	0.25	2.47	6.38	3.45	1.93	0.37
	Std	0.59	2.63	2.87	7.14	0.74	4.53	8.42	6.00	4.87	1.43
	Med	0.00	1.00	0.00	0.00	0.00	0.00	0.4 0	0.00	0.00	0.00
Mann Whitney		-4.16**	-4.57**	-5.32**	-4.76**	-3.15**	-6.20**	-5.85**	-6.69**	-3.34**	-2.62**

**p<0.01 *p<0.05

Subj = Subjective; Obj = Objective; prsnt = present; futr = future; artcl=article; vrb=verb; axvrb=auxiliary verb, numb. = number

Table 17: Linguistic attributes of serviceable posts and its sub-types with statistical significance using Mann-whitney and Kruskal Wallis test.