Community-in-the-loop: Creating Artificial Process Intelligence for Co-production of City Service

YE WANG*, University of Missouri - Kansas City, USA
SRICHAKRADHAR REDDY NAGIREDDY, University of Missouri - Kansas City, USA
CHARAN TEJ THOTA, University of Missouri - Kansas City, USA
DUY H. HO, University of Missouri - Kansas City, USA
YUGYUNG LEE, University of Missouri - Kansas City, USA

Communities have first-hand knowledge about community issues. This study aims to improve the efficiency of social-technical problem-solving by proposing the concept of "artificial process intelligence," based on the theories of socio-technical decision-making. The technical challenges addressed were channeling the communication between the internal-facing and external-facing 311 categorizations. Accordingly, deep learning models were trained on data from Kansas City's 311 system: (1) Bidirectional Encoder Representations from Transformers (BERT) based classification models that can predict the internal-facing 311 service categories and the city departments that handle the issue; (2) the Balanced Latent Dirichlet Allocation (LDA) and BERT clustering (BLBC) model that inductively summarizes residents' complaints and maps the main themes to the internal-facing 311 service categories; (3) a regression time series model that can predict response and completion time. Our case study demonstrated that these models could provide the information needed for reciprocal communication, city service planning, and community envisioning. Future studies should explore interface design like a chatbot and conduct more research on the acceptance and diffusion of AI-assisted 311 systems.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; *Interaction paradigms*; **Natural language interfaces**.

Additional Key Words and Phrases: artificial process intelligence; community-in-the-loop; co-production; 311 calls; Bidirectional Encoder Representations from Transformers; Latent Dirichlet Allocation

ACM Reference Format:
Ye Wang, Srichakradhar Reddy Nagireddy, Charan Tej Thota, Duy H. Ho, and Yugyung Lee. 2022. Community-in-the-loop: Creating Artificial Process Intelligence for Co-production of City Service. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 285 (November 2022), 21 pages. https://doi.org/10.1145/3555176

1 INTRODUCTION

Many cities in the United States have adopted the 311 system, including phone calls and Twitter, as a platform for non-emergency residents' requests and delivering city services. This includes all kinds of community issues that require city services, ranging from the trash in neighborhoods, dead animals, potholes, to parking violations, water leaking, and other criteria. 311 call records

Authors' addresses: Ye Wang, WanYe@umkc.edu, University of Missouri - Kansas City, Kansas City, Missouri, USA, 64110; Srichakradhar Reddy Nagireddy, University of Missouri - Kansas City, Kansas City, USA, snp8b@mail.umkc.edu; Charan Tej Thota, University of Missouri - Kansas City, Kansas City, USA, thota.charantej@gmail.com; Duy H. Ho, University of Missouri - Kansas City, Kansas City, USA, dhh3hb@mail.umkc.edu; Yugyung Lee, University of Missouri - Kansas City, Kansas City, USA, leeyu@umkc.edu.

2573-0142/2022/11-ART285 $15.00
https://doi.org/10.1145/3555176

include time, location, and text data. They document thousands of calls from residents to inform the city management of problems that need city services. Then, the city management assigns problems to workgroups that deliver the service. As a result, the large amounts of data from the 311 system have been examined to guide future planning of resource deployment or to hold public officials accountable for deficiencies [35].

Some studies conceptualize this co-production process as resident-as-consumer and government-as-provider. However, such a consumer-provider model cannot convincingly explain the socio-spatial disparities in 311 complaint behavior, particularly in Kansas City, MO [12]. As pointed out by e-government research, involvement and engagement of all constituents is key to co-producing efficient and fair public service [12]. To address this community involvement and engagement challenge, this study proposes a community-in-the-loop approach that leverages “artificial process intelligence" to create a support system for the 311 system.

2 BACKGROUND

2.1 Motivation

The first wave of science studies in the 1950s and 1960s has proven that science is not able to solve social problems (community problems are essentially social) without two-way communication and a collaborative process of co-creation with the communities [7]. Moreover, communities have long-time first-hand experience with community issues, which should receive equal or comparable weight via delegation (like elected public servants) or representation (like opinion polls and surveys) [7]. Thus, community engagement has become an essential component of the second and third waves of science studies.

Thus, today’s challenge is not to argue for the importance of community-in-the-loop but to create a mechanism to address the methodological question of “how" to bring communities in the loop. For example, a recent study funded by the UK’s Economic and Social Research Council spent three years examining how autism research could become more participatory, i.e., relevant and transformative to the autism community [9]. The findings revealed that the major obstacles were a lack of supportive components [1] for communities to get heard and involved and a mechanism that allowed multi-directional communication among all stakeholders for constructive dialogues [1, 9].

Applying this socio-technical decision-making perspective, scholars argued that engineering models and urban design should make an effort to understand community concerns, aspirations, and adaptation ideas and explore and evaluate solutions in terms of their responsiveness to the local context and economic viability [26]. The case study on urban flood resilience in Australia documented the use of visualization techniques to facilitate the process of community visioning participated by community stakeholders [26]. Community visioning gave insight into the requirements and desires of the community, which subsequently led engineering modeling and analysis, as well as urban design [26]. Thus, the first and foremost step of a community-in-the-loop approach is to leverage technologies like Machine Learning and Deep Learning to more effectively and efficiently listen to community inputs and understand human insights. This “listening" process is essential to human-centric computing.

To address these challenges, we propose the concept of “process intelligence" to highlight the kind of artificial intelligence required to facilitate the listening process. The theoretical foundation of “artificial process intelligence" is the concept of “process experts." It was proposed in Treem, and Barley [32] and developed by Barley, Treem, and Leonardi [1]. It refers to personnel “structurally” at the intersection between domain experts with knowledge about operational, curational, evaluative,

and representational processes. For instance, the nurse station is strategically situated at the intersection of physicians or 311 operators at the intersection of city service departments.

Ideally, "process experts" should be able to understand "situated knowledge of the social processes operating in a specific social context, including the knowledge of accessing, synthesizing, and evaluating volumes of information, and presenting information in a way that is meaningful to all stakeholders [1]. Furthermore, "process expertise" requires intelligence and wisdom based upon extensive knowledge (big data) of "who the stakeholders are," "what the types of expertise/issues are involved," "when (coordinating multiple agendas and timelines)," "where (what are the environmental, contextual, and situational factors)," "why (multiple causes and consequences from various perspectives)," and "how (evaluating different solutions)." Although efforts have been made to grow process expertise from humans, there are insurmountable obstacles of high individual commitment (both time and effort) to acquire extensive knowledge and institutional undervaluation of process expertise in comparison to domain expertise [1].

2.2 Problem Statement: Bridging the Gap between Internal and External Categorization of 311 Complaints

The challenge for automation is comprehending citizens' reports of events, classifying them into 311 service categories, identifying city departments (workgroups) that provide services, and communicating with the residents about the key information regarding the problem and the service delivery. Figure 1 illustrates the communication processes of the 311 call system: (1) data collection; (2) modeling; and (3) informing. This project aims at the "modeling" phase that generates "process intelligence." The specific challenges facing "modeling" are (1) to address the challenges of semantically arbitrary classes in the existing internal-facing 311 department and the service category categorization; (2) to bridge the gap between the internal and the external categorization of complaints.

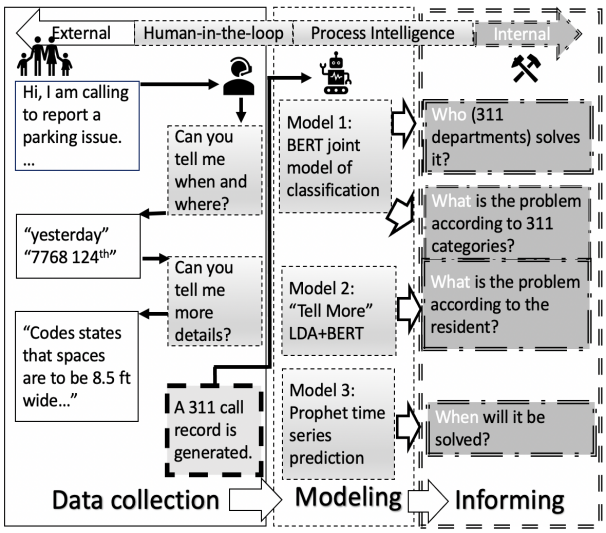

Fig. 1. 311 Use Case and the Overall Framework

The call records of 311 systems involve navigating and connecting two categorization systems: (1) an external-facing system, from which the operator collects information about the problem from residents who call in; (2) an internal-facing categorization system to navigate the city service

departments for problem solving. This dual-system presents two major challenges to human operators and an AI support system.

First, the internal-facing categorization needs to categorize events into semantically arbitrary classes. This means that some of the 311 categories that the model needs to handle do not have a focused semantic theme. Instead, they are standard practices or legacy categories inherited from older systems. This is not an uncommon situation in real-world classification tasks. For example, 311 service categories contain 17 categories: "city facilities," "sidewalks," "government," "streets," "lights," etc. The categorization is not mutually exclusive. It does not follow any hierarchical categorizations (events and sub-events). The 311 department categories include "policy department," "finance," "general service," "Northland," "South," etc. Some of these categories are functional, and others are geographical. These internal-facing categories become so complicated that it is hard for human operators to efficiently sort them out so that city service can be delivered on time. An AI support system should be able to offer assistance to facilitate internal categorization.

The first problem leads to the second problem: The semantically arbitrary categorizations for internal communication are not designed for efficient external communication with the most important stakeholders of the city management, i.e., the residents. Residents' calls are unstructured text or voice data, which do not follow any internal-facing categorization. An AI support system needs to comprehend the problem reported by the resident and then use the information to communicate with the residents regarding key milestones of the problem-solving process, for example, who (which city department) will solve the problem, and when the service will be delivered to where. The current human-operated 311 system is not sufficiently resourced to follow up with residents reciprocally.

2.3 The Proposed Solution

Accordingly, the proposed solution is designed to answer the same set of questions that a process expert can answer: What community problem happened, when and where it happened, and who was involved. Adopting this community-in-the-loop approach, we presented a case study of Kansas City, MO's 311 system. Machine learning and Deep Learning models were employed to create a human-centric AI system for more efficient and effective two-way communication. It includes three models that can communicate with each other:

- *Model 1. "Who handles the problem" and "What happened according to the internal-facing categorization:"* We designed supervised learning models using the Bidirectional Encoder Representations from Transformers (BERT) model of classification to conduct service category and department category classifications of 311 calls.
- *Model 2. "Can you tell more" and "What happened from the resident's perspective:"* We proposed the Balanced Latent Dirichlet Allocation (LDA) and BERT CLustering (BLBC) model that inductively categorizes unstructured inputs from residents and maps them onto the internal-facing categorization.
- *Model 3. "When and where the problem will be solved:"* We designed a time-series forecast model that predicts when the requested service will be delivered to the requested location.

Figure 1 shows the communication functions of these three models in an episode of a 311 call.

3 RELATED WORK

3.1 Event Detection

To create such an AI support system, event detection (ED) and event classification are at the center of transforming 311 call records into intelligence. The existing ED approaches mostly solve two problems: (1) detecting an event from a large amount of text data, like Tweets. For example, machine

learning models were created to improve stock market forecasting to detect significant events from Tweets and align the event with the timeline [18]; (2) classifying an event into a category. An example is to classify an event of hiring someone into the category of "personnel: start-position" [13].

There are two major approaches to event detection (ED) tasks: topic modeling and event trigger detection. Topic modeling is a text dimension reduction technique [18]. It is used to detect "event topics" and "sub-event detection" in, for example, tweets [18]. Even trigger-based approaches reply on reference datasets like ACE2005, to detect and classify event trigger words. ACE2005 annotation guidelines define event trigger words as "the word that most clearly expresses its occurrence." An event word can be the main verb of the sentence or a noun, an adverb, etc.. An example of an event trigger word is, for example, "hire" in "Emily was *hired* to take over as chief executive."

Event trigger-based classification uses deep learning models based upon feature engineering that selects event trigger words [13]. Recent studies also investigated methods of combing event trigger words with contextual words to enhance the performance [20]. However, the event trigger-based classification often uses supervised machine learning, which causes difficulty in performing classification on unseen classes [13]. Recognizing the limitation of event trigger-based classification, Ngo et al. [20] applied few-shot learning event detection.

These existing approaches of ED are not immediately applicable to our case. As aforementioned, the communication process of 311 involves a dual-categorization system. It is necessary to combine the supervised classification approach with the unsupervised topic modeling approach.

3.2 Deep Learning and Natural Language Processing

Recently, significant advances have been made in Natural Language Processing and Deep Learning, such as ELMO [23], GPT family [4, 24, 25], BERT [8], RoBERTa [16], XLNet [36]. In particular, BERT has been extensively explored in conjunction with various NLP models to achieve state-of-the-art performance. The pretrained uncased BERT is introduced in Devlin et al.[8]. It generates representations from unlabelled text data by jointly conditioning on both left and right contexts (bi-directional) in all layers. It can be fine-tuned with just one additional output layer for various tasks.

Peinelt et al.[22]'s topic-informed BERT-based model (tBERT) combined topic representation of a sentence from LDA topic models with the sentence pair vector C ($C = BERT(S_1, S_2) \in R^d$) of $BERT_{BASE}$ where d specifies the hidden internal size of BERT, i.e., 768 for $BERT_{BASE}$. S_1 is a sentence with length N, and S_2 is another sentence with length M. Both are the uncased version of $BERT_{BASE}$, which does not differentiate between english and English. Employing $BERT_{BASE}$, Peinelt et al.[22]'s iBERT with LDA topics produced accurate and stable performance across a range of benchmark datasets of semantic similarity prediction.

In contrast to Peinelt et al.[22]'s model-driven approach, Venkataram et al. [34]'s experimentation with LDA and BERT was largely data-driven. Answering the call from the White House, they aimed at exploiting the COVID-19 Open Research Dataset consisting of more than 29,000 machine-readable articles on COVID. Their TopiQAL was an "interpretable, unsupervised, generic and fused" ML and DL architecture for COVID-related question-answering. They used LDA and BioBERT, and their unique contribution was the hierarchical inference that matches user query sentences with LDA topic distribution of abstracts with a probability threshold = 0.2, followed by the same process on paragraphs in body text. Two levels of topic model filtering supplied chosen topics to the BERT extractive summarizer for Q&A. BioBERT was BERT adapted for the biomedical domain. [14].

3.3 Time Series Forecasting

The traditional univariate forecasting techniques predict future values of time series based on their historical values [10]. However, amid the criticism of the black-box nature of the artificial neural network, recent successes of recurrent neural network (RNN) models have shown great potential [10]. For example, Long Short Term Memory (LSTM), which is an artificial RNN, was trained and tested on datasets of COVID transmission in Canada and Italy to predict future outbreaks [6].

Some studies found that deep learning models like LSTM produced more accurate and clear patterns and predictions than mathematical and statistical models [6]. They can learn to identify non-linear patterns and explore latent relationships without any prior [29]. Smyl [29] mixed an exponential smoothing (ES) model with LSTM into one common framework that combines the strength of statistic models and neural networks. This hybrid forecasting method used exponential smoothing for deseasonalizing and normalizing the series and LSTM for extrapolating the series [29]. Livieris et al. [17] created a CNN-LSTM model to predict the price of gold. This approach uses CNN to preprocess data and screen out noises, and an LSTM layer is stacked on top of it to perform forecasting. Their first CNN-LSTM model without a fully connected layer performs well on regression tasks, like predicting prices. Their second CNN-LSTM model with a fully connected layer performs well on classification tasks like predicting the gold movement. Niu et al. [21] combined two-stage feature selection, convolutional LSTM, Gated Recurrent Unit (GRU), and an error correction model to predict the financial market.

Sirignano et al. [28] proposed a time series price prediction with a 4-layer perceptron model for price changes in Limit Order Books. Neil et al. [19] proposed an LSTM architecture for asynchronous series detection by tacking learning dependencies of various frequencies in the time series. Borovykh et al. [3] proposed a predictive model with convolutional neural networks for conditional time series. These related studies show that real-world data often require different techniques. Therefore, assessing models on real-world data is needed to find the best practices for performing time-series forecasting.

4 THE PROPOSED MODELS

Our use case requires (1) classifying an event into existing internal-facing 311 service categories and city departments. We propose to use BERT as an embedding method, combined with supervised learning. (2) To detect the major community problems reported by citizens, we propose to use the Balanced LDA+BERT Clustering model to map the results of LDA topic modeling to the internal-facing 311 service categories. (3) A time series forecasting model to predict the delivery of the service. We will compare statistical models with DL models. The major contribution of this paper is creating “process intelligence" by applying multiple models and connecting the “knowledge" from each to close the loop for the community-in-the-loop approach. See Figure 1 for the overall framework.

4.1 Model 1: BERT-based 311 Call Classification

There are two classification tasks given a 311 call record: (1) the 311 service department and (2) the 311 service categories. As we intended to use models from this study to create a questions answering system, we considered event classification as a natural language understanding (NLU) problem and designed BERT models of classification [37].

BERT is Bidirectional Encoder Representation from Transformers that relies entirely on self-attention instead of sequence-aligned recurrent neural networks (RNNs) or convolutions. It consists of multiple bidirectional transformer encoder layers [33]. Each layer, surrounded by a residual connection, has a multi-head self-attention mechanism, followed by a position-wise fully connected

feed-forward network. An attention function can map a query and a set of key-value pairs to an output. The output is a weighted sum of the values, and the weight assigned to each value is a compatibility function of the query with the corresponding key. The attention weights are calculated by Equation 1: the three inputs are Q queries, K keys, and V values; and the output is the softmax of standard dot-product attention, QK^T of Q and K (K^T represents the transpose of matrix K) with a scaling factor of $\sqrt{d_k}$, where d_k is the dimension of the key, ensuring the value of the dot product does not grow too large.

$$Attention(Q,\ K,\ V) = softmax(\frac{\mathrm{QK}^T}{\sqrt{d_k}})V \tag{1}$$

Our classification model includes two parts: a BERT encoder that encodes 311 records/descriptions and a classification decoder that classifies a 311 call record into a 311 category (BERT-C) or a service department (BERT-D). The two classification tasks in problem and solution domains are trained separately.

We used the BERT encoder transformer with an added layer of classification decoder. The encoding of a 311 service description is described in Equation 2. x_i is the representation of each token and t is the number of the tokens, and h_i is the contextual semantic representation embedding of a token. Thus, $H = (h_1, \ldots, h_t)$, the encoder outputs are the semantic representations of each 311 record.

$$H = BERT(x_1, \ldots, x_t) \tag{2}$$

Given an input token sequence $X = (x_1, \ldots x_t)$, the output of the BERT encoder is $H = (h_1, \ldots, h_t)$, and h_i is the averaged output from the multi-headed transformer blocks given as token's contextual semantic representation embedding.

The hidden representation $H \in R^{|X|h}$ is obtained by H = BERT(X), where $|X|$ is the length of the input sequence $X = (x_1, \ldots x_t)$ and h is the size of the hidden dimension. Then, H is passed to a dense layer $W \in R^{h|V|}$, followed by softmax, as described in Equation 3.

The classification decoder uses sentence semantic representation H to predict the class label y^c:

$$y^c = softmax(WH + b) \tag{3}$$

y^c can be the 311 categories (BERT-C) or service departments (BERT-D). Softmax is an activation function that converts a vector of numbers into probabilities within the range of [0, 1]:

$$\sigma\,(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{4}$$

$\vec{z}$ is input vectors. e^{z_i} is the standard exponential function for input vectors. K is the number of classes. e^{z_j} is the standard exponential function for output vectors.

4.2 Model 2: Balanced LDA+BERT Clustering (BLBC)

As stated in the second problem, we need a model to categorize residents' complaints inductively and connect them with the internal-facing 311 service categories. The internal-facing 311 service categories are sometimes semantically arbitrary and thus hard for the external audience to understand.

To solve this problem, we propose a new topic modeling, called Balanced LDA+BERT Clustering (BLBC), by combining LDA [2] and BERT [8]: LDA [2] is first used to detect topic per document probabilities, which is then combined with BERT [8] sentence embedding through an autoencoder. Finally, the latent representations from the encoder are entered into a clustering algorithm to categorically cluster documents.

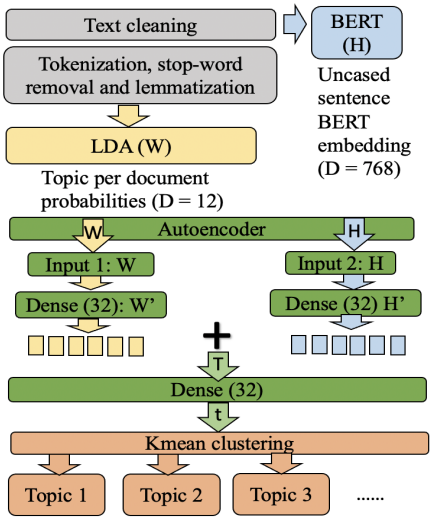

Fig. 2. Architecture of Balanced LDA+BERT Clustering (BLBC)

Pre-trained language models: As our model is based on pre-trained BERT, we briefly describe the BERT model here. The BERT document vector H is generated using Equation 1 and Equation 2. The self-attention mechanism is described in Equation 1. Given an input document D, the uncased BERT model outputs the semantic representations of the document H, as described in Equation 2.

We intend to combine the strengths of LDA and BERT. The performance of the LDA topic model could be influenced by the number of documents, the length of documents, and the number of topics [30]. Shorter texts, like 311 call records, may suffer from poor performance due to their length. This can be attributed to the LDA model's random drawing of the document-topic, and topic-word proportion vectors [30]. While the length of 311 call records may undermine the performance of LDA, the coherence flow between sentences with each document may introduce opportunities to improve the performance of the topic model. Unlike tweets and short texts from social media, 311 call records are short. Still, each document may have a more coherent sentence flow since they are human-generated records of calls from residents about a specific instance. Here is an example of a 311 call record:

"Citizen reports improper parking space striping. Codes state that spaces are 8.5 ft wide, but the spaces are only 8ft wide. The handicapped spaces are only 7.5 ft wide but are supposed to be 8.5 ft wide. The problem is likely to be present throughout the parking garage. Additionally, there should be handicapped parking signage in front of each stall, but there are none, just the logo on the ground."

This record has a clear semantic connection from sentence to sentence, and the entire record, due to high sentence-level coherence, is semantically focused. The same observation was made by Li et al. [15] on texts from Wikipedia. They proposed a bi-Directional Recurrent Attentional Topic Model (bi-RATM) for document embedding to capture sentence-to-sentence flow, and their model achieved state-of-the-art performance [15]. In the same spirit, our topic model uses BERT sentence-level embedding to overcome the possible less-than-optimal performance of LDA on shorter texts (see Figure 2).

LDA topic vectors: The LDA document vector is generated by Latent Dirichlet Allocation (LDA). It is a generative probabilistic model based on three primary structures, including words, topics, and documents [2]. The documents are represented as a random mixture of latent topics, and each topic is characterized by a distribution over words. Given an input corpus D ($d \in D$) with V unique

words and M documents, each document d contains a sequence of N words $d = \{W_1, W_2, \ldots, W_N\}$, $n \in \{1, 2, \ldots, N\}$. Given a topic number K, $k \in \{1, 2, \ldots, K\}$, the generative process will generate documents based upon per-document topic distribution and per-topic word distribution and optimize probabilities. α is the per-document topic distribution. It is a matrix where each row is a document, and each column is a topic. It indicates the likelihood that a document contains topic Z_k. β is the per-topic word distribution. This matrix has rows to represent topics and columns words, indicating how likely a topic Z_k, k contains word W_n. θ_d is a multinomial distribution of documents drawn from a Dirichlet distribution with the parameter α. φ_k, k is a multinomial distribution of words in a topic drawn from a Dirichlet distribution with the parameter β. For each word position n, select a hidden topic Z_n from the multinomial distribution parameterized by θ_d. And, then select W_n from φ_{Z_n}.

$$P(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\varphi}; \alpha, \beta) = \prod_{i=1}^{K} P(\varphi_i; \beta) \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t} \mid \theta_j) P(W_{j,t} \mid \varphi_{Z_{j,t}}) \tag{5}$$

Autoencoder: The LDA document vector W is defined below:

$$W_i = TopicModel([T_1, \ldots, T_N]) \in R^t \tag{6}$$

T_i is the probability of the document belonging to the i-th topic. N is the number of topics.

W is Input 1 of the encoder E. A fully connected neural network is employed to learn W', vector representations of W. W' has the same dimension d as H'. w_1 is the weights, and b_1 is the bias:

$$W' = ReLU(w_1 \times W + b_1) \tag{7}$$

The BERT document vector H has a dimension of $d' = 768$. It is entered into the encoder E as Input 2. The fully connected neural network learns H', vector representations of H. H' has the same dimension d as W'. w_2 is the weights, and b_2 is the bias:

$$H' = ReLU(w_2 \times H + b_2) \tag{8}$$

W' and H' are concatenated into a single document vector T.

$$T = W'H' \tag{9}$$

A full-connected NN learns latent vector representations of T. w_3 is the weights, and b_3 is the bias:

$$TopicVectors(T) = softmax(w_3 t + b_3) \tag{10}$$

The decoder D mirrors the dimensions and layers of the encoder E.

Clustering: The vector representation of each document is the hidden state of the autoencoder. E is the encoder, and D is a document:

$$t_i = E(D_i) \tag{11}$$

Clusters of the documents, where each cluster indicates a topic category, are generated using a clustering algorithm like K-means (see Figure 2).

4.3 Model 3: Time series forecasting.

The Prophet model was used to predict the estimated responding time for a specific 311 service call using the 311 service data from the past ten years. The Prophet model is a Generalized Additive Model (GAM) [31]. The Prophet model includes three major components: $g(t)$ is the trend, $s(t)$ is seasonality, $h(t)$ is holidays, and ϵ_t is the error term. They are summed up to perform a forecast:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{12}$$

$g(t)$ is modeled using the piece-wise logistic growth model as follows: The trend changes in the growth model have been adjusted by explicitly defining change points, where the growth rate is allowed to change. For the given S change points at times s_j, $j = 1, \ldots, S$, a vector of rate adjustments is defined as $\delta \in R^S$, where δ_j is the change in rate that occurs at time s_j. The rate at any time t is then the base rate k, plus all of the adjustments up to that point: $k + Pj : t > s_j \delta_j$. A vector is defined as $a(t) \in \{0, 1\}^S$ and the rate at time t is then $k + a(t)^T \delta$. When the rate k is adjusted, the offset parameter m is also adjusted to connect the endpoints of the segments. The correct adjustment γ at change point j is defined as follows:

$$g(t) = \frac{C(t)}{1 + exp((k + a(t)^T \delta)(t - (m + a(t)^T \gamma)))} \tag{13}$$

with $C(t)$ is time-varying capacity, k the growth rate, and m an offset parameter. Also, seasonality models have been defined as periodic functions of t using the Fourier series, considering periodic effects to fit and forecast these effects.

We have also explored several deep learning models, including LSTM and CNN models, for time series forecasting. Surprisingly, our case study showed the Prophet forecasting model had superior prediction compared to the deep learning models to be discussed later, suggesting city service delivery is influenced by seasonality and holidays.

5 KCMO 311 CASE STUDY

5.1 311 Data

The domain of our study is the open dataset 311 of the Kansas City metropolitan area (from 2015 to 2020). Some of the key fields from the data are: Departments of 311 services, categories of 311 services, descriptions of the incidents, geo-location coordinates of the incidents, and dates of the 311 service requested and delivered.

The 311 calls can be reported through multiple communication channels such as phone calls, email, Web, or social media. Most of them (about 97%) were reported by the three channels: phone (approximately 70%), Web (about 20%), and email (about 6.7%).

5.2 Model 1: The KCMO 311 Service Category and Department Predictive Model

Data: The KCMO 311 service data [11] (shown in Table 1) is split into 80-20 train-validation ratio to train and evaluate the model's performance. As seen in Figure 3, the total numbers of the internal-facing 311 service categories and departments are 17 (not including the "other" and "no data available" categories) and 15, respectively.

Table 1. KCMO 311 Service Request Dataset

Department#	Category#	Training#	Testing#	Total
15	17	112,412	28,103	140,515

Training: To train the 311 BERT classification models of categories and departments of the 311 service requests, the 311 data was split between training and validation in the ratio of 80-20. As a result, the 311 BERT classification model has been trained in 10 epochs. The accuracy of the 311 service category and department classification was approximately 95.5% and 96.15%.
Evaluation with Baselines: We used two measures to measure the performance of the classification algorithms: (i) The loss function (Eq. 14) used for deep neural networks is cross-entropy to measure the difference between the predicted labels and the proper labels; (ii) Accuracy Eq 15.

$$Cross - entropy = -\sum_{i=1}^{m}\sum_{j=1}^{n} y_{i,j}\log(p_{i,j}) \tag{14}$$

where $y_{i,j}$ denotes the true value $p_{i,j}$ denotes the probability predicted by the model of a sample i belonging to class j, m is the number of the classes, and n is the size of a training set.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{15}$$

where TP is true positive, FN is false negative, FP is false positive, and TN is true negative.

The evaluation of the proposed predictive models has been conducted in comparison to other machine learning algorithms, including Support Vector Machines (SVM), Decision Tree, Naive Bayes, and K-means Clustering.
311 Call Classification Results The class-wise accuracy for the 311 service category and department classification are shown in Figure 3. The Recreation and Park category has the lowest accuracy (82.61%) in the 311 category classification. The Northland department is the least accurate (67.50%) among the departments. As shown in Table 2, the BERT classification models performed the best. SVM was the second-best among the five different algorithms (92.96% and 93.77%), and Decision Tree was the worst (44.66% and 62.95%).

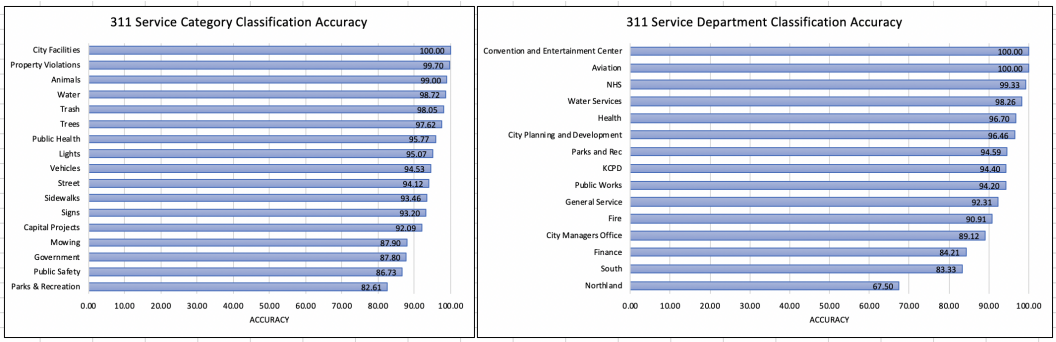

Fig. 3. Class-wide Classification Accuracy: (a) 311 Service Category (b) 311 Service Department

Discussion about 311 Call Classification Models Figure 4(a) shows that categories with lower accuracy (e.g., “Parks & Recreation") were largely due to data imbalance issues. Categories with higher frequencies tended to have the highest accuracy, while those with lower frequencies had the lowest accuracy. Although “Property Violations" enjoyed a 99.7% accuracy, the miscategorized cases suggested some overlapping with “Street" and “Trash."

Figure 4(b) shows departments with higher frequencies had higher prediction accuracy. Among the lowest four departments, “Finance," and “South" had fewer than 60 cases. The confusion occurred

Table 2. 311 Service Category and Department Prediction Accuracy

Model Name	**Service**	**Department**
Naive Bayes	81.46%	85.89%
K-means Clustering	82.91%	85.08%
SVM	92.96%	93.77%
Decision Tree	44.66%	62.95%
BERT Transformer	**95.50%**	**96.15%**

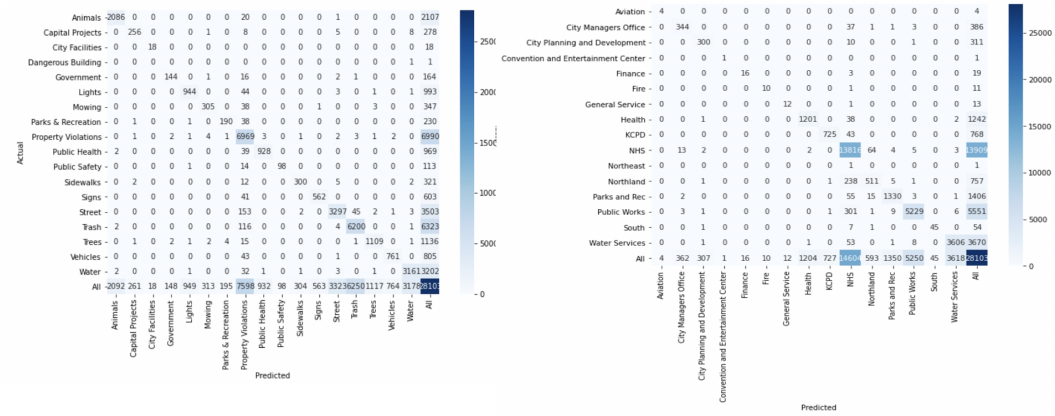

Fig. 4. Confusion Matrix: (a) Service Category Classification (b) Department Classification

mostly between “Northland" and “City Managers Office" and “NHS (National Honor Society)." The mixed-up can be attributed to the crossover between function-based department classification (like National Honor Society) and jurisdiction-based department classification.

5.3 Model 2: BLBC Clustering Modeling

Data: All 311 text records were used. The goal of Model 2 was inductively to summarize the main themes of residents’ complaints and then analyze the relationship between the main themes/topics of complaints with the internal-facing 311 service categories. We achieved this goal by (1) summarizing the main themes/topics of the complaints by using LDA topic modeling; (2) clustering documents into topic categories by using the Balanced LDA+BERT Clustering (BLBC) model; and (3) visualizing the relationship between themes/topics and internal-facing 311 service categories.

Training the LDA Topic Model: The preprocessing includes text cleaning, tokenization, stop-word removal, and lemmatization. To ensure optimal results, coherent scores were calculated. Figure 5 shows the coherence values across a range of 8 to 25 topics, and 12 topics had the highest topic coherence (0.5512). 17 topics (the existing 311 services have 17 categories) had a coherence score of 0.5350. In the experiments, we demonstrated that the quality of the LDA topic modeling had a significant impact on the BLBC clustering models.

Results of the LDA Topic Model: Table 3 showed the top 10 words of the 12 topics from the LDA model. We conducted a topic-by-topic qualitative comparison of the 12-topic and the 17-topic LDA models since the internal-facing 311 service categories have 17 categories. Both the 12-topic and the 17-topic LDA models identified “traffic," “animals," “limbs and lights," “water leak," “bulky items," “vehicles," “trees and potholes," “houses," and “property maintenance." Topic 10 “Trash" of the 12-topic model was split into three topics in the 17-topic model. The 17-topic model had two

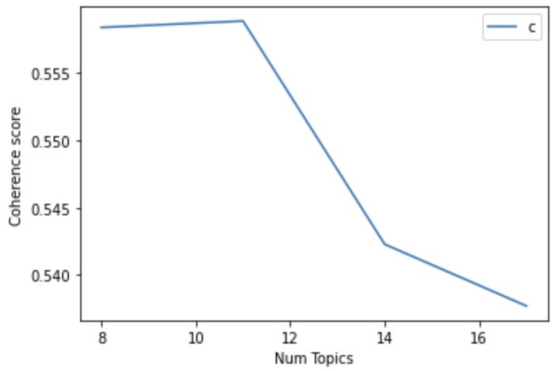

Fig. 5. Coherence for Optimal Topic Modeling

Table 3. Top 12 Topics and Topic Terms

ID	Topic	Topic Terms (10)
Topic 1	Traffic	traffic, corner, hydrant, lane, hole, fire, street, north, road, south
Topic 2	Animals	dog, black, taker, white, brown, pit, large, small, loose, flat
Topic 3	Limbs and lights	limb, dead, street, light, note, cartegraph, replace, pole, possible, soon
Topic 4	Case referred	refer, please, case, see, note, work, still, service, closed, missing
Topic 5	Trucks	truck, get, city, last, week, time, disabled, put, state, know
Topic 6	Water leak	water, leak, meter, home, pressure, low, coming, pipeline, street, state
Topic 7	Bulky items	property, item, bulky, recycle, leaking, neighborhood, dumping, sewer, maintenance, appointment
Topic 8	Vehicles	vehicle, car, parked, street, sign, plate, abandoned, parking, front, month
Topic 9	Trees and potholes	tree, street, city, need, side, pothole, sidewalk, large, right, removed
Topic 10	Trash	trash, missed, bag, violation, sticker, picked, recycling, collected, block, time
Topic 11	Houses	house, front, tire, open, door, building, damaged, side, home, vacant
Topic 12	Property maintenance	yard, property, grass, weed, lot, back, need, cut, tree, debris

topic categories that were hard to interpret. The top 10 words of these unclear themes were: 1) get, people, need, time, cover, American, ice, concern, state, area; 2) meter, refer, see, please, note, traveler, turned, construction, flat, driveway. Within the 12 topics, Topic 5 Trucks appeared to be a bit challenging to interpret. This comparison showed that the 12-topic model summarized the themes better than the 17-topic for a human to interpret.

Training of BLBC: Text cleaning was conducted, and sentence BERT embeddings were generated using pre-trained uncased BERT.

The BLBC model was trained for 5 epochs. The encoder was saved and used to generate hidden layer vector representations.

Evaluation of BLBC: The objective function of BLBC was log-cosh. The log-cosh loss function is a regression loss function that behaves similarly to the mean squared loss but is robust to outliers. It is the logarithm of the hyperbolic cosine of the prediction error. Formally:

$$L(y,\ y^p) = \sum_{i=1}^{n} log(cosh(y_i^p - \ y_i)) \tag{16}$$

We conducted four experiments to compare BLBC with other models. First, we compared BLBC with the baseline model. The baseline model replicated the existing LDA+BERT model: The LDA

Table 4. Silhouette Scores (SS) of Topic Models

Models	Dimension of W	#Clusters	W'/H' Ratio	SS Score
BLBC (Ours)	12	12	**32:32**	**0.3623**
ILBC-1	12	12	12:52	0.2985
ILBC-2	17	12	17:47	0.2405
ILBC-3	17	17	17:47	0.2321
Baseline [27]	12	12	N/A	0.2745

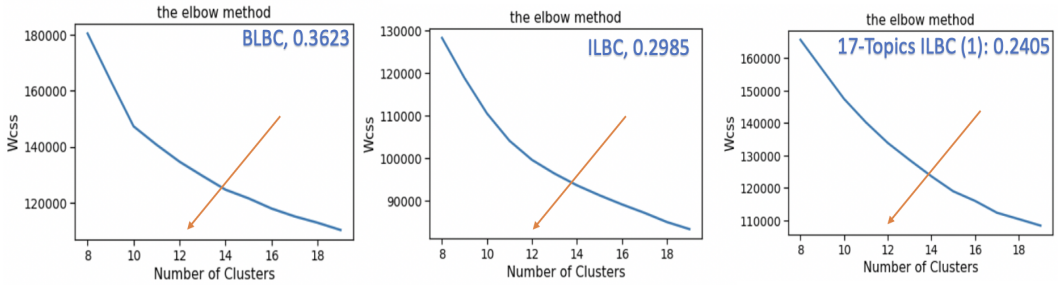

Fig. 6. Elbow Charts: (a) Balanced LDA+BERT Clustering (BLBC), (b) Imbalanced LDA+BERT Clustering (ILBC), and (c) 17-Topics ILBC(1)

topic vector W (from the 12-topic LDA model) was concatenated with the BERT document vector H. This joint vector T was entered into a shallow autoencoder with one dense layer that reduced the dimension of T to 32. The hidden layer vector t ($d = 32$) was entered into a cluster model, to cluster the documents into 12 categories. The second comparison was to experiment with different dimensions of W' and H': 1) $d'_W = 12$, $d'_H = 52$, and the number of clusters $N = 12$ (ILBC-1 in Figure 6); 2)$d'_W = 17$, $d'_H = 47$, $N = 12$ (ILBC-2 in Figure 6). The third comparison was to compare different LDA topic vectors: 17-topic vector vs. 12-topic vector: $d'_W = 17$, $d'_H = 47$, $N = 17$ (ILBC-3 in Figure 6).

All models were trained for 5 epochs. The latent representations of the proposed model BLBC, the original LDA+BERT model (baseline), ILBC-1, ILBC-2, and ILBC-3 were entered into the same K-means clustering algorithm.

We evaluated the clustering models using the Silhouette score and the elbow method. The Silhouette score measures how coherence a document is to its own cluster compared to the other clusters. It has a range of $[-1, 1]$, with 1 indicating high cohesion within the cluster and high separation from the other clusters. Thus, it is a better measure than visualization when dealing with higher dimension clustering. The elbow method is a visual method to identify the optimal number of clusters by plotting WCSS (Within-Cluster Sum of Square).

Results of the BLBC Model: Our BLBC model is different from the existing LDA+BERT model since it balances the LDA topic and BERT document vectors. We demonstrated the importance of balancing the LDA topic vector and the BERT document embedding by experimenting with the length of W', and H'. A balanced length of the vectors performed the best and produced the highest Silhouette score (see Table 4). We demonstrated the impact of the LDA topic model (coherence) on the clustering model by using the 17 topic vector in the same autoencoder (see 17-Topics ILBC-1 and 17-Topics ILBC-2 in Figure 6). 12 topics were of higher quality than 17 topics and thus produced better results of clustering.

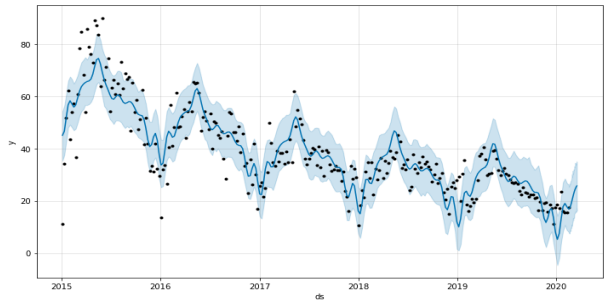

Fig. 7. Time Series Analysis

Discussion of the BLBC Model: Tables 5 - 7 show this relationship. The cross-tabulation tables include the 17 service categories (including the “other" and the “data not available" categories). The counts of the documents per topic/theme were based upon the membership from the BLBC clustering model. The 12 main topics/themes of complaints were concentrated on the top categories of the internal-facing 311 service categories. “Trash," “Water Leak," and “Animals" were three major problems that both categorization systems agreed upon. The “Traffic" problem involved both the lights/signals/signs service and the street/sidewalks service. The “Limbs and Lights" problem was largely a public safety concern. The problems of “Houses" and “Property Maintenance" involved services of property violations and street/sidewalks. Table 5(a) showed that a majority of the complaints in the “other" category belonged to the themes/topics of “Trees and Potholes," “Property Maintenance," and “Bulky Items." The “other" category is often created to group items that are hard for human operators to put them into a category.

5.4 Model 3: KCMO 311 Time Series Forecasting

Data: The time series forecasting was conducted on all the data available from 2015 to 2020. The results are shown in Figure 7.
Training: To further drill down upon the trends of the case resolution, we have analyzed the same data based upon the category of the 311 service requests. Due to lack of data, only cases since 2008 were entered into the analysis. Therefore, the category-wise trends have been shown in Figure 8.
Evaluation: Figure 9 shows the evaluation of time series results on a 70-30 split of data.

The time series model for the 311 service data, which is a nonlinear regression based model for 311 call response times, was evaluated: (1) MAE (Mean Average Error) (defined in Eq. 17) is the measure of the difference between predicted versus observed, (2) MAPE (Mean Average Percentage Error) (defined in Eq. 18) is a measure of prediction accuracy of the forecasting (loss function for regression in machine learning).

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{17}$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \tag{18}$$

Results of the Time-series Mode Figure 9 shows the MAE and MAPE scores as the performance of the 311-time series model. The best classes of the response time prediction are *Animal* and *Tracy* with 1.23% and 6.27% of its MAE. The worst class accuracy is *City Facility* with 29.3% of MAE. In terms of MAPE, the best class is *Lights* with 7.45% of MAPE, and the worst type is *Tracy*, with 46.93% of MAPE. Thus, the overall accuracy for the response time estimation is about 70%. The

Table 5. Topics and Dominant 311 Service Categories

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Topic 1	1049	5602	319	46	984	7501	1334	79	2460	563
Topic 2	20748	38	73	177	30	45	139	41	27	10
Topic 3	3581	4545	466	3	9055	1649	2120	10	116	306
Topic 4	735	1070	2384	674	871	1919	2821	156	4357	1043
Topic 5	1025	1274	2560	3455	1498	3688	8692	972	2442	1242
Topic 6	3337	36	1173	715	84	903	492	130	22682	357
Topic 7	59	72	5887	79	352	430	8603	204	252	3475
Topic 8	218	3495	2059	31	155	1235	748	9864	1161	306
Topic 9	650	767	3090	53	3216	13120	2349	79	1289	5212
Topic 10	428	18	1124	15	38	120	46120	6	44	269
Topic 11	1462	138	6503	558	433	861	2520	149	562	547
Topic 12	765	423	7616	346	2602	2154	5942	538	2378	4977

Topic 1: Traffic; Topic 2: Animals; Topic 3: Limbs & Lights; Topic 4: Case Referred; Topic 5: Trucks; Topic 6: Water Leak; Topic 7: Bulky Items; Topic 8: Vehicles; Topic 9: Trees, Potholes, Limbs, Lights; Topic 10: Trash; Topic 11: Houses; Topic 12: Property Maintenance
Topic Terms: T1 Animals; T2. Lights, Signals, Signs; T3. Property Violations; T4. Public Health; T5. Public Safety; T6. Street, Sidewalks; T7. Trash; T8. Vehicles; T10. Water; T11. Other

Table 6. Topics and Dominant 311 Service Categories

Traffic	Citizen calling to report the traffic lights going north and south are stuck on red and going east and west the lights are stuck on green. Red Bridge and Hickman Mills Dr.
Animals	The citizen is reporting a dog bite that happened at this address. The incident happened on 06/15/16 between 6:30p and 7:30p The bit a human victim and another dog. The attacking dog is white with black.
Limbs & Lights	Citizen is reporting twol street lights out. The two pole numbers are SDM1010 and SDM1011.
Case Referred	The citizen is requesting a callback in regards to case number 2015072174. The note from 7/9/2015 states that a letter was mailed to the citizen on 6/5/2015 and the citizen has not received the letter.
Trucks	No snow plow or salt truck has been down the street. I know deadends are last but it said to wait 36 hours and it has been longer than that. Thank You.
Water Leak	Citizen called to report water leak. Water leak is located in the street around this address. Water is clear and odorless. Water is trickling from both sides of the pavement.
Bulky Items	Citizen s requesting bulky appointment but address is not in the scheduler.
Vechicles	Citizen is calling to report an abandoned white Chevy van sitting in front of this location. The van has been sitting on the street for months.
(1) Trees, Potholes, (2) Limbs, Lights	(1) The caller is reporting two ROW trees located at the curb be removed due to the roots of the tree are buckling his sidewalks. (2) Citizen reports city oak trees located on the right of way along W 69th Ter and Ward parkway need to be trimmed.
Trash	Citizen called to report had 2 bags of trash out said that the trash truck collected 1 bag and left the other, also did not collect neighbors trash.
Houses	The citizen is calling to report that the house is open to entry. The doors and the windows are all off.
Property Maintenance	The citizen reports the grass and weeds are overgrown in the front and back yard. There is brush in the yard. The yard has not been mowed in over 2 months now.

Prophet forecasting model showed superior performance when compared to the deep learning models (Table 8).

Discussion of the Time-series Mode The time-series model shows a steadily decreasing trend overtimes, which suggests an improvement in the performance of the city service. The cases that were used to take around 80 days to resolve in 2015 were solved in under 20 hours in 2020. The most significant improvement was “Animals", “Government", “Mowing", “Vehicles", and “Water". “Capital Projects," “Public Safety," “Sidewalks," “Signs," and “Trees" have a slower and flatter downward trend.

Table 7. Topics and Minor 311 Service Categories

	T1	T2	T3	T4	T5	T6	T7	T8	T9
Topic 1	100	381	14	335	3	4	0	0	124
Topic 2	1	28	2	5	2	0	0	0	0
Topic 3	10	58	3	33	3	1	0	0	1
Topic 4	43	232	41	207	10	10	21	3	18
Topic 5	42	407	51	300	13	2	25	59	36
Topic 6	28	62	14	17	5	2	1	1	1
Topic 7	12	56	8	40	19	9	1	0	0
Topic 8	34	70	6	94	9	0	4	0	151
Topic 9	102	308	15	275	20	12	0	0	16
Topic 10	10	42	0	18	6	0	0	0	0
Topic 11	23	132	12	42	6	1	4	0	2
Topic 12	88	878	23	140	35	16	7	1	7

Topic 1: Traffic; Topic 2: Animals; Topic 3: Limbs & Lights; Topic 4: Case Referred; Topic 5: Trucks; Topic 6: Water Leak; Topic 7: Bulky Items; Topic 8: Vehicles; Topic 9: Trees, Potholes, Limbs, Lights; Topic 10: Trash; Topic 11: Houses; Topic 12: Property Maintenance
T1. Parking; T2. Parks Recreation; T3. City Facilities; T4. Legal; T5. NA; T6. Maintenance; T7. Neighborhood; T8. Noise; T9. Traffic

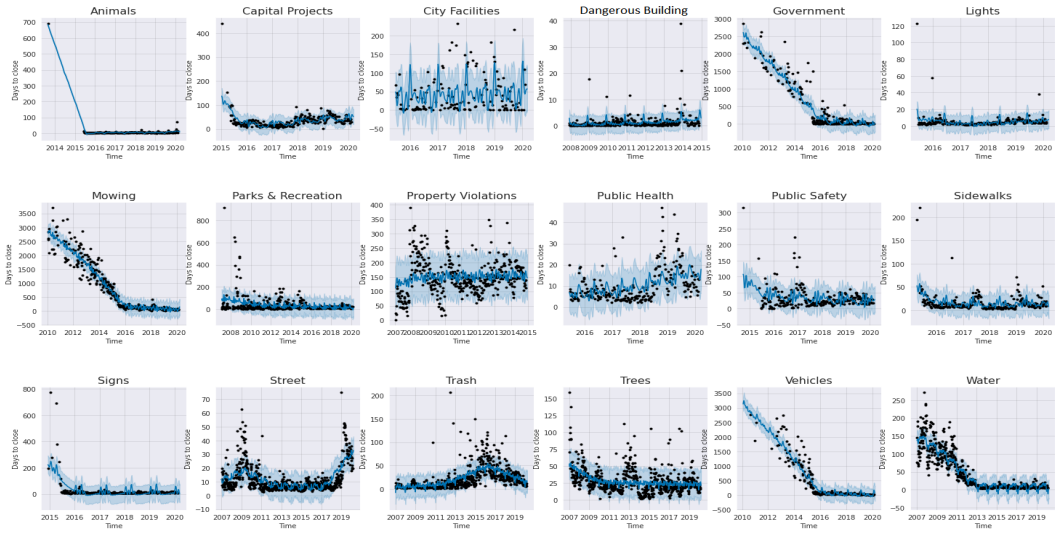

Fig. 8. 311 Service Category Wise Time Prediction

Table 8. 311 Time Series Predictive Model Accuracy

Model Name	**MAE**	**MAPE**
LSTM	9.71	105.63
LSTM + Window	7.81	79.33
LSTM + Time Steps	5.70	65.87
LSTM + Memory	7.44	84.50
Stacked LSTM + Memory	7.84	93.62
DCNN (SeriesNet)	14.20	210.32
Prophet (Proposed)	**3.45**	**14.38**

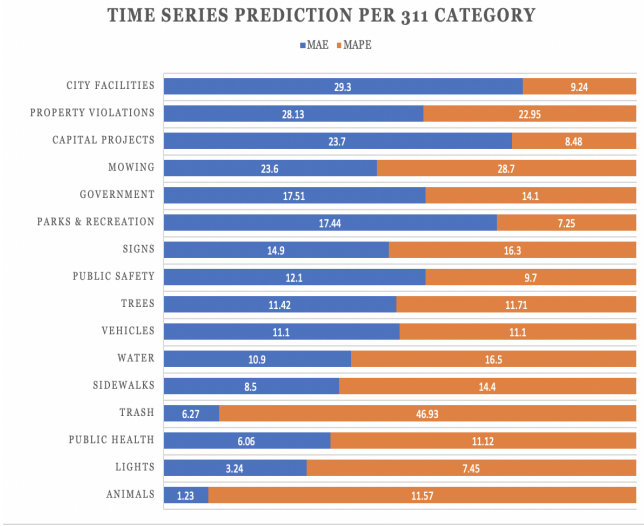

Fig. 9. 311 Call Time Series Validation

“City Facilities," “Dangerous Building," and “Lights" exhibit strong seasonality, without a significant downward or upward trend. “Property Violations," “Public Health," “Street," and “Trash" show highly idiosyncratic patterns; the outbreaks of the problems rather than seasonality influences service time. These are the most problematic areas of 311 service delivery.

6 DISCUSSION

This study aims to address the “how" question of bringing communities into the loop of city service co-production. The methods were three Deep Learning/Machine Learning models that generated “process intelligence." The three models serve different communication purposes. The BERT classification models facilitate internal communication with departments and workgroups. They can predict the service categories and the service departments. The LDA topic model, the Balanced LDA+BERT clustering (BLBC) model, and the time-series prediction model facilitate external communication with the residents. The classification models and the BLBC model can map onto each other via the relationship between service categories and main topics/themes of complaints. The time series forecasting model provides an estimated time of service delivery. Connecting the intelligence from each model is key to “artificial process intelligence." Our following discussion will focus on the implication of this study on creating a support system for communication and planning and the long-term goal of community-in-the-loop city-residents co-production.

First, the “artificial process intelligence" can support human operators by offering recommendations on service departments, service categories, and predicted delivery time in real-time. This is an important step toward closing the loop of two-way communication. As shown in Figure 10, the models can help human operators improve the efficiency of the conversation, and provide critical information of the problem-solving process, reassuring residents that the service will be provided. Second, being applied as a support system, the models can give suggestions to human experts on classifying the calls, communicating with the residents, and following up the calls. For example, a large number of call records ($N = 18,307$) were categorized as “other.” These call records did not have a service category. They thus were not included in the training of the classification model. Our analysis showed that the majority of the “other” category belonged to the topics of “Bulky

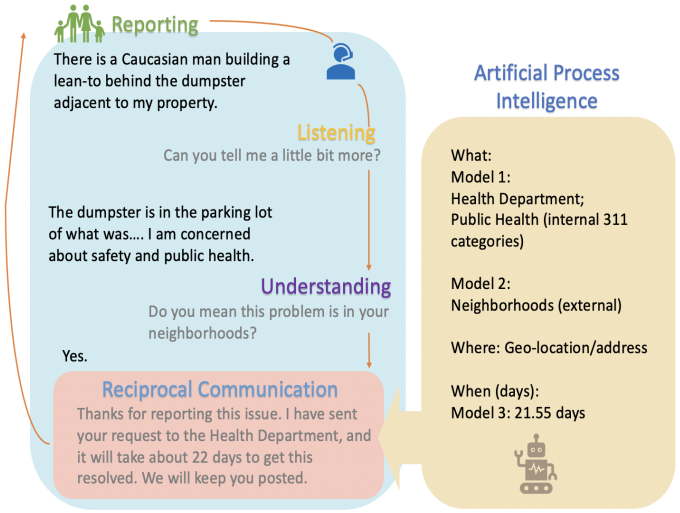

Fig. 10. Closing the Loop for Community-in-the-loop

Items," "Trees and Potholes," and "Property Maintenance." The BLBC model can map these topics onto internal-facing 311 service categories. Accordingly, the BLBC model can suggest a problem type, and related service categories given the content of a call. As a result, the AI assistant can reduce the number of cases in the "other" category and potentially improve the efficiency of human categorization and service delivery.

Based on these models, future studies can develop dialogue agents, who handle 311 calls during after-hours (by adding automatic geo-location detection). The inductive topic models (LDA and BLBC) create the foundation for a conversation agent to comprehend an incoming call, respond to residents' inquiries, and route the problem to the corresponding service category and the service department.

Second, the "process intelligence" can support the planning of city management. The LDA topic model and the BLBC model can supplement existing human analysis by providing summaries of major problems and discovering the prevalence of these problems. For example, Table 5 and Table 6 show that the major problems in Kansas City are trash, trees/potholes/limbs/lights (combining Topic 9 and Topic 3), water leak, and property maintenance/houses (combining Topic 11 and Topic 12). However, the internal-facing service categories ranked the problems differently. The top service categories were trash, water, animals, and streets/sidewalks. Property violations (N=33,254) ranked fifth in the internal-facing service categories, in contrast to N = 42,891 cases in the topic categories of property maintenance and houses. The mapping between property violations from the internal categorization and property maintenance/houses from our topic model shows that property violations missed a certain amount of trash issues relating to properties. Additionally, many cases ($N = 4,977$) relating to property maintenance were put into the "other" category. In other words, property maintenance/houses may be a bigger issue than the internal categorization may show. For example, the internal-facing service categorization shows that animals were a bigger problem than property violations—this contrast echos with some of the local concerns voiced by the communities. A recent news report by KCUR pointed out that property vacancy and abandoned houses may be a bigger problem as the Land Bank database may show [5].

7 LIMITATIONS

The long-term goal of “artificial process intelligence" is to bring communities, policy-makers, urban designers, and scientists together to solve community issues. While this study has made advancements on the methodological question of “how to bring communities into the loop," technologies may not be sufficient to address the socio-spatial disparities in co-production without human-machine collaboration. Models provide insights and intelligence. “Artificial process intelligence" makes it easier for residents and city managers to form a shared vision of city services and community development. Human efforts, however, are needed to disseminate and communicate the knowledge from the models, so stakeholders will trust the technology, trust each other, and form a consensus. Future studies may need to conduct more research on the acceptance of AI-assisted 311 systems and the diffusion of technological innovation among residents and city management.

In terms of computer-human interaction design, this proof-of-concept study created models without creating an interface for the application. Thus, collecting human users' feedback at the application level is an important next step. The impact of data imbalance issues was more significant than expected. For less frequent 311 categories, the accuracy of the models' prediction dropped. In addition, overlapping or possible hierarchies in complaint categorization were not adequately addressed in this study. This may potentially reduce the efficiency of the models at the application level.

ACKNOWLEDGMENTS

The co-author, Yugyung Lee, would like to acknowledge the partial support of the NSF, USA Grant No. 1747751, 1935076, and 1951971.

REFERENCES

[1] William C Barley, Jeffrey W Treem, and Paul M Leonardi. 2020. Experts at coordination: Examining the performance, production, and value of process expertise. *Journal of Communication* 70, 1 (2020), 60–89.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[3] Anastasia Borovykh, Sander Bohte, and Cornelis W Oosterlee. 2017. Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv:1703.04691* (2017).

[4] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165* (2020).

[5] Celisa Calacal and Emily Wolf. 2021. Vacant lots, absentee owners, little accountability. What's going on with the Kansas City Land Bank? https://github.com/Stveshawn/contextual_topic_identification. [Online; accessed 8-August-2022].

[6] Vinay Kumar Reddy Chimmula and Lei Zhang. 2020. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals* 135 (2020), 109864.

[7] Harry M Collins and Robert Evans. 2002. The third wave of science studies: Studies of expertise and experience. *Social studies of science* 32, 2 (2002), 235–296.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[9] Sue Fletcher-Watson, Jon Adams, Kabie Brook, Tony Charman, Laura Crane, James Cusack, Susan Leekam, Damian Milton, Jeremy R Parr, and Elizabeth Pellicano. 2019. Making the future together: Shaping autism research through meaningful participation. *Autism* 23, 4 (2019), 943–953.

[10] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. 2021. Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting* 37, 1 (2021), 388–427.

[11] kcmo. 2013. KCMO 311. https://www.kcmo.gov/city-hall/departments/neighborhoods-housing-services. [Online; accessed 13-September-2020].

[12] Constantine E Kontokosta and Boyeong Hong. 2021. Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions. *Sustainable Cities and Society* 64 (2021), 102503.

[13] Viet Dac Lai, Franck Dernoncourt, and Thien Huu Nguyen. 2020. Extensively matching for few-shot learning event detection. *arXiv preprint arXiv:2006.10093* (2020).

[14] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36, 4 (2020), 1234–1240.
[15] Shuangyin Li, Yu Zhang, and Rong Pan. 2020. Bi-directional recurrent attentional topic model. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 14, 6 (2020), 1–30.
[16] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).
[17] Ioannis E Livieris, Emmanuel Pintelas, and Panagiotis Pintelas. 2020. A CNN–LSTM model for gold price time-series forecasting. *Neural computing and applications* 32, 23 (2020), 17351–17360.
[18] Elliot Maître, Zakaria Chemli, Max Chevalier, Bernard Dousset, Jean-Philippe Gitto, and Olivier Teste. 2020. Event detection and time series alignment to improve stock market forecasting. In *Joint Conference of the Information Retrieval Communities in Europe (CIRCLE 2020)*, Vol. 2621. CEUR-WS. org, 1–5.
[19] Daniel Neil, Michael Pfeiffer, and Shih-Chii Liu. 2016. Phased lstm: Accelerating recurrent network training for long or event-based sequences. In *Advances in neural information processing systems*. 3882–3890.
[20] Nghia Trung Ngo, Tuan Ngo Nguyen, and Thien Huu Nguyen. 2020. Learning to select important context words for event detection. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 756–768.
[21] Tong Niu, Jianzhou Wang, Haiyan Lu, Wendong Yang, and Pei Du. 2020. Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. *Expert Systems with Applications* 148 (2020), 113237.
[22] Nicole Peinelt, Dong Nguyen, and Maria Liakata. 2020. tBERT: Topic models and BERT joining forces for semantic similarity detection. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 7047–7055.
[23] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365* (2018).
[24] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.
[25] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.
[26] BC Rogers, Nigel Bertram, Berry Gersonius, Alex Gunn, Roland Löwe, Catherine Murphy, Rutger Pasman, Mohanasundar Radhakrishnan, Christian Urich, THF Wong, et al. 2020. An interdisciplinary and catchment approach to enhancing urban flood resilience: a Melbourne case. *Philosophical Transactions of the Royal Society A* 378, 2168 (2020), 20190201.
[27] Stve Shawn. 2019. Contextual topic identification for Steam reviews. https://github.com/Stveshawn/contextual_topic_identification. [Online; accessed 8-August-2022].
[28] Justin A Sirignano. 2019. Deep learning for limit order books. *Quantitative Finance* 19, 4 (2019), 549–570.
[29] Slawek Smyl. 2020. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting* 36, 1 (2020), 75–85.
[30] Jian Tang, Zhaoshi Meng, Xuanlong Nguyen, Qiaozhu Mei, and Ming Zhang. 2014. Understanding the limiting factors of topic modeling via posterior contraction analysis. In *International Conference on Machine Learning*. PMLR, 190–198.
[31] Sean J Taylor and Benjamin Letham. 2018. Forecasting at scale. *The American Statistician* 72, 1 (2018), 37–45.
[32] JEFFREY W TREEM and WILLIAM C BARLEY. 2016. 11 Explaining the (De) valuation. *Expertise, Communication, and Organizing* (2016), 213.
[33] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.
[34] Hamsa Shwetha Venkataram, Chris A Mattmann, and Scott Penberthy. 2020. TopiQAL: Topic-aware Question Answering using Scalable Domain-specific Supercomputers. In *2020 IEEE/ACM Fourth Workshop on Deep Learning on Supercomputers (DLS)*. IEEE, 48–55.
[35] Wei-Ning Wu. 2020. Determinants of Citizen-Generated Data in a Smart City: Analysis of Open 311 User Behavior. *Sustainable Cities and Society* (2020), 102167.
[36] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems*. 5753–5763.
[37] Zhichang Zhang, Zhenwen Zhang, Haoyuan Chen, and Zhiman Zhang. 2019. A joint learning framework with bert for spoken language understanding. *IEEE Access* 7 (2019), 168849–168858.

Received April 2021; revised November 2021; accepted March 2022