A Call Center Model for Online Mental Health Support

Tim Rens de Boer
CWI
Amsterdam, Netherlands

Saskia Mérelle
113 Suicide Prevention
Amsterdam, Netherlands

Sandjai Bhulai
Vrije Universiteit
Amsterdam, Netherlands

Rob van der Mei
CWI and Vrije Universiteit
Amsterdam, Netherlands

***Abstract*—Helplines for mental healthcare differ from other call centers in various aspects. Many agents are volunteers, the conversations are often more complex and emotional, and many helplines use a triage system. In this paper, we first propose a call center model that includes the specifics of online mental health helplines, including features such as a triage system for chats and service times consisting of a warm-up, conversation, and wrap-up cool-down periods. The model is validated using a trace-driven simulation based on real-life (anonymous) data provided by 113 Suicide Prevention. The results show that the model can simulate the waiting-time performance of the helpline accurately. Second, we focus on forecasting the number of chats and telephone calls. Our results show that (S)ARIMA models trained on historical data perform better than other models in the case of short-term forecasting (five weeks or less ahead), while using linear regression works best for long-term forecasts (longer than five weeks).**

***Index Terms*—call center models; queueing; mental health; helplines; data analytics; forecasting**

I. Introduction

There are many forms of mental health, with many countries having one or multiple helplines. Examples of mental health helplines in the Netherlands are 113 Suicide Prevention [1], the helpline for help-seekers with suicidal thoughts, the listen helpline (Dutch: luisterlijn) [2], and the Kindertelefoon [3] for children. Recently, mental-support helplines have received much attention due to the increase in call volumes related to the (partial) lockdown to combat the spreading of corona [4], [5]. This paper focuses on call center modeling for the suicide prevention helplines. However, the results can also be used for modeling the waiting-time performance of other mental health helplines.

Suicide is a worldwide health problem. In 2020, on average, five persons died each day by suicide in the Netherlands alone. Suicide is a leading cause of death among adolescents [6]; worldwide, more than 700,000 people die from suicide every year [7]. In many countries, people struggling with suicidal thoughts can contact a helpline to get support to prevent and reduce suicidal thoughts [8]. In the Netherlands, persons with suicidal thoughts can contact 113, either by telephone or by chat [9], and are helped by volunteers and professionals. It is crucial that help seekers are answered swiftly. Therefore, it is important that adequate staffing is present to answer telephone calls and chat requests, minimizing the waiting times and abandonments (i.e., sudden termination of ongoing calls or chats while waiting). In order to calculate proper staffing levels (e.g., [10]), a good understanding of the processes of call center system for mental health is required.

Mental health helplines differ in various aspects from classical call centers. First, the subjects of conversations are mental health concerns, e.g., loneliness, substance use [11]–[13]. Second, agents often have to handle complex conversations with the patient-in-need, and may themselves require support during or after a difficult conversation [14]. Therefore, an important aspect is that agents often need some time to cool down after emotionally difficult conversations. Third, when a chat or telephone call enters the system, it has to be determined which agent is best capable of handling the call, which results in a warm-up time before the call is taken into service. Lastly, chat conversations that enter the system first go through a triage phase. The inclusion of triage plays an important role, and functions as a filter [15] to chat requests, checking whether these help seekers are at the right helpline or might require emergency care. The triagist can also estimate the conversation's difficulty to best match an agent with enough experience to handle the chat. So, the model for such a mental health helpline needs to meet the following requirements: (1) the possibility of abandonments, (2) a service time consisting of multiple phases, (3) a warm-up and cool-down time, (4) inclusion of chats and telephone calls, and (5) triage.

Classic call center models were considered for modeling this helpline, see [16] for an excellent overview on queuing models for call centers. The Erlang-C model is proven to be inaccurate when modeling call centers with abandonments [17]. The Erlang-A model includes abandonments [18], but does not include multiple skills, triage, and warm-up times. Multi-skill call center models, such as an N-configuration [16], were also considered. The N-configuration model includes two types of arrivals, which in this case would be telephone and chat, and assumes that new chat arrivals can be picked up before triage. However, in the mental health helpline of [1], it is crucial that chats first go through triage. The importance of modeling triage is already shown in the emergency domain [19], [20]. Therefore, in this paper, we modify the N-configuration model in such a way that it includes the specifics of mental health helplines.

This paper introduces a new queuing model to include the possibility of triage. The model is built based on anonymized call and chat data, made available by [1]. The records used cannot be traced back to individual callers, as only timestamps

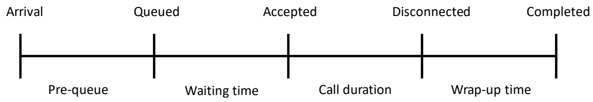

Fig. 1. Timeline of a call from the perspective of the caller.

and durations are used. The anonymous data is also used to validate the model and determine different patterns that can be helpful for forecasting demand.

This paper is organized as follows. Section II describes different patterns found in the data. In Section III, the proposed model is described. In Section IV, the model is validated, using a combination of data and trace-driven simulation. Next, Section V explains how the demand call volumes can be predicted based on historical data. Finally, in Section VI, the conclusion and discussion are given.

II. Data Analysis

The data for this research is provided by 113 Suicide Prevention [1], throughout referred to as '113'. Its mission is to prevent suicides and break the taboo surrounding suicide. Help seekers struggling with suicidal thoughts can contact 113 24/7 by either chat or telephone. Apart from that, 113 also provides online therapy, self-tests, and self-help courses [21].

Call data is provided over the period 2016-2021 and consists of around 250,000 chats and 175,000 telephone calls. Each call or chat has the following elements: (1) the arrival time, (2) the time entering the queue, (3) the time of acceptance by an agent, (4) the disconnection time, and (5) the completion time. Each call also has a contact id and an *initial* contact id, which may be different if a call is a forwarded telephone call or chat.

The time between the arrival of a call and the time entering the queue is the time spent by the caller in a menu. This phase does not require resources from the helpline and, therefore, falls outside the scope of this paper. The *waiting time* is defined as the time between entering the queue and the time that an agent accepts the call. The *call duration* where the agent and the help seeker are actually connected is defined as the time between acceptance of a call and the disconnection time. Finally, after each call, the agent has to fill in a wrap-up form, which is the time between disconnection and completion. A visual representation of this timeline is given in Fig. 1.

Recall that there are two options for help seekers to contact the helpline: via *telephone* or *chat*. These two contact types are mostly handled by the same type of agents but differ in some important aspects. Traditionally, there used to be more chats than telephone calls. However, the difference has diminished in recent years, and the numbers are comparable. However, the chat and telephone calls do follow different patterns, which are shown in Figs. 2 to 5. Figs. 2 and 3 show the week patterns; in both cases, the weekends see a lower number of calls. We also observe that most telephone calls arrive between 12:00 and 20:00, while chats have a clear peak at 20:00. Both telephone and chat call arrivals show a decrease during the night and early morning.

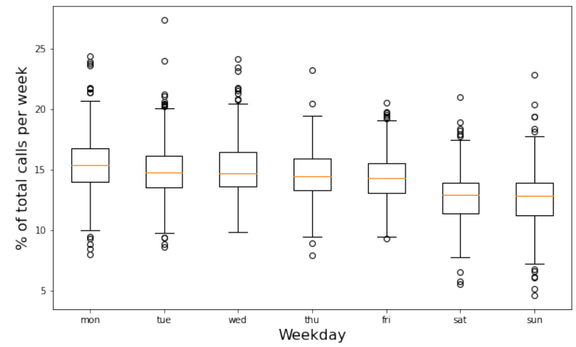

Fig. 2. Weekly pattern of incoming telephone calls.

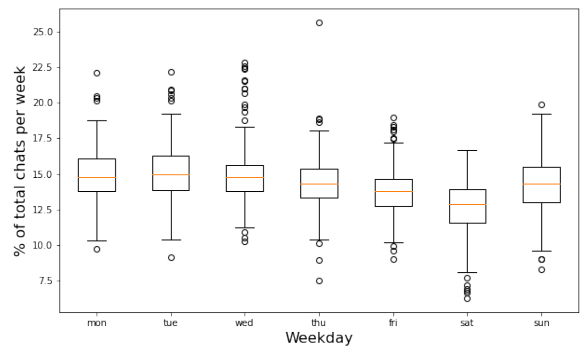

Fig. 3. Weekly pattern of incoming chats.

Only part of the incoming chats after triage gets sent to another agent. The percentage of chats that an agent handles after triage is around 50% during day shifts, but this differs over the day. During the night shifts, fewer chats are forwarded to an agent. The different nature of night conversations may cause this. The triage plays an important role in filtering chats, as seen by the percentage of chats that get through triage. Chats are filtered out due to various reasons. For example, the chatter may be at the wrong helpline and or is identified as a

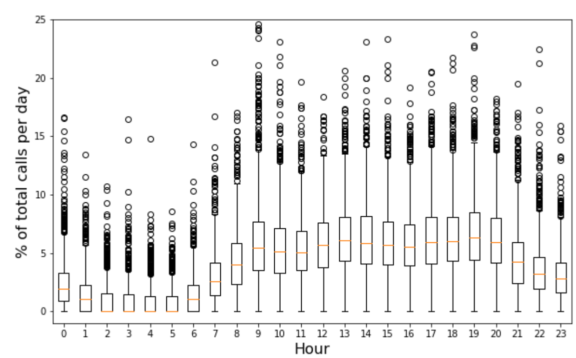

Fig. 4. Daily pattern of incoming telephone calls.

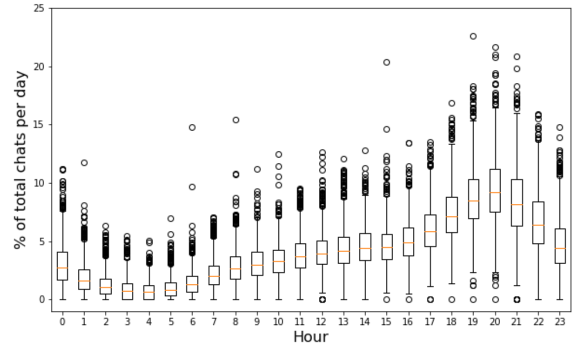

Fig. 5. Daily pattern of incoming chats.

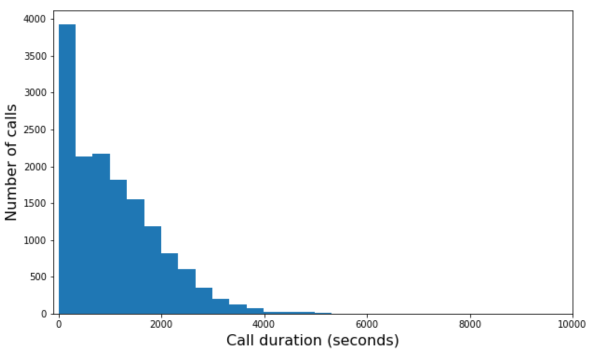

Fig. 6. Histogram of telephone call durations.

prank chatter [22].

In the data, we distinguish three service time distributions: for the duration of (1) telephone calls, (2) chats *during* triage, and (3) chats *after* triage. The empirical distribution of phone call durations can be found in Fig. 6. On average, a telephone call takes around 18 minutes. The duration of chats in triage can be seen in Fig. 7, and the duration of chat conversations after triage can be seen in Fig. 8. As can be seen, the chats that have gone through triage tend to take much longer than the chats during triage: on average, chats in triage take 19 minutes versus 34 minutes after triage.

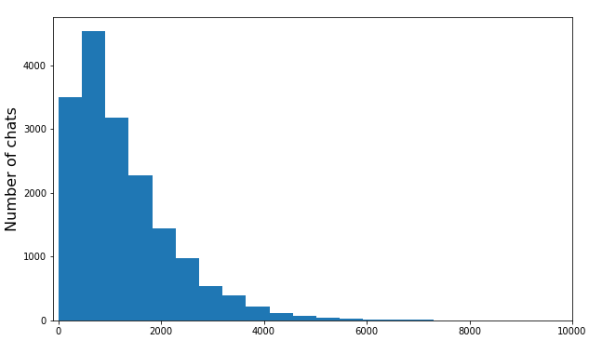

Fig. 7. Histogram of chat triage durations.

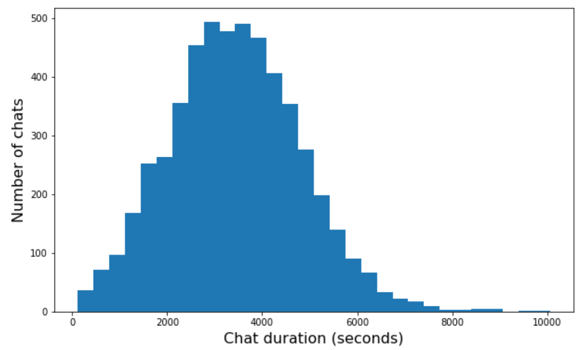

Fig. 8. Histogram of durations of chat conversations after triage.

Remark 1: The effect of experience of agents
Mental health helplines often use a mix of paid professionals and volunteers, where an agent's responsibilities depend on experience and other factors. Analysis of our data pointed out that there seems to be no significant difference in service time distribution between different volunteers and professionals. There might be a difference, but this could be obscured due to many other factors, such as experienced agents handling the more complex conversations. Based on this observation, the call duration distributions are assumed to be the same for all agents.

Remark 2: The effect of media events on arrivals
The possible effects of media events concerning suicide were also studied. If shown to have a significant and long-lasting effect, these media events could be used to help predict demand. To analyze this, multiple events were considered, ranging from national and international suicides to political news concerning suicide. The effect of suicides of celebrities on suicides is well-documented. It has been found that during and after the suicide of a celebrity, the number of suicides increases [23]. Rather surprisingly, we observed that most events do *not* have a large or long-lasting effect on the number of call arrivals, often having no effect or only a short-term effect for 1 or 2 days.

III. Model description

There are two types of calls: (1) *chat calls*, and (2) *telephone calls*. Incoming chat calls first enter the triage system that consists of c_{triage} triage agents. When a chat call finds a triage agent available, it enters the triage immediately. Each triage agent can handle n_{triage} simultaneously without perceivable slow-down; this way, there are $c_{\text{triage}} n_{\text{triage}}$ 'triage slots' for chat calls. The service time of a chat call in triage, denoted B_{chat}, is the convolution of four random variables $B_{\text{chat}} = B_{\text{warm-up}} + B_{\text{conversation}} + B_{\text{wrap\_up}} + B_{\text{cool-down}}$, drawn from some probability distribution, which can be estimated from the data. A visual representation of the service time can be seen in Fig. 9. Incoming chat calls that find all $c_{\text{triage}} n_{\text{triage}}$ triage slots occupied, take place in an infinite-sized queue and are handled on a FCFS basis.

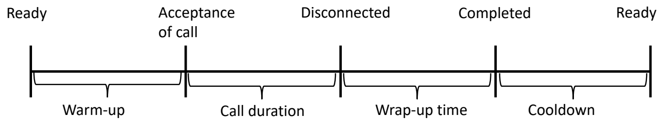

Fig. 9. Timeline of call from the perspective of the agent.

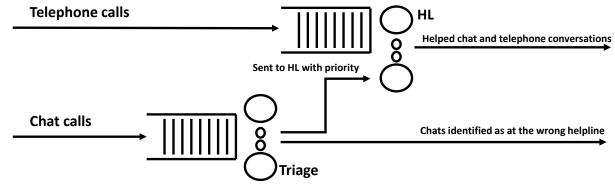

Fig. 10. Illustration of the model.

A key aspect of the model is the inclusion of *impatience*, i.e., the maximum amount of time that a help seeker is willing to wait before he abandons the system. The impatience of a help seeker who enters the system via a chat is modeled as an independent sample from some probability distribution with mean $\mu_{\text{chat-impatience}}$. After service completion at the triage center, a chat is either sent through to the helpline (HL) for assistance (with probability $p_{\text{sent-through}}$), or the chat leaves the system. Note that during the warm-up period $B_{\text{warm-up}}$, the agent is busy, but the help seeker is not yet answered. Therefore, help seekers may abandon the queue during that period.

Different from chat calls, incoming telephone calls do *not* go through a triage phase and arrive directly at the HL. This is the core part of the system where most of the service processing occurs. The HL is equipped with c_{HL} agents, each of which can handle both chat calls and telephone calls, not more than one call at a time. When a telephone call finds an HL-agent available, he enters service immediately. If the telephone call arrives and no agent is available, the call enters an infinite-sized queue that is handled on an FCFS basis.

The HL processes both telephone calls and forwarded chat calls (i.e., those that have passed through the triage phase). Here, *chat calls have non-preemptive priority over telephone calls*. Thus, when an HL-agent becomes idle, he first checks whether there is a chat call pending (while keeping a triage slot occupied), and, if so, starts to service the longest-waiting chat call. If no chat call is pending, the agent checks if telephone calls are pending.

Similar to the modeling of chat sessions, the duration of phone calls also consists of four subsequent independent phases: (1) warm-up, (2) conversation, (3) wrap-up, and (4) cool-down, where each phase has its own probability distribution.

The impatience of help seekers via telephone is modeled as a sample from some probability distribution that can be obtained from the data. Calls abandon the queue if their waiting time exceeds the impatience. When service is completed, the calls exit the system. Note that agents are busy during the warm-up period, similar to abandonments at the triage, but the help seeker does not perceive this and can therefore still abandon the queue during the warm-up phase. See Fig. 10 for an illustration of the model.

IV. Model validation

The model is validated using trace-driven simulation. For this, the following data is used for each call: (1) arrival time, (2) conversation duration, and (3) wrap-up duration. However, for the simulation, the following missing values have to be filled in: (a) the conversation and wrap-up duration of calls that were unanswered, (b) impatience of help seekers, and (c) warm-up and cool-down durations. The missing values of conversation and wrap-up durations were filled in using hot-deck-imputation [24]. This method samples from the known values to fill in the missing values. The distinction was made between the different types of conversations: telephone, chats during triage, and chats after triage. The impatience of help seekers is mainly unknown due to the limited availability. Only a small percentage of telephone calls and chats were unanswered. Warm-up and cool-down durations were not present in the data. Therefore, the missing values of impatience, warm-up, and cool-down were all drawn from exponential distributions. The parameters were estimated using expert opinions from paid professionals and volunteers. Impatience of chat conversation is determined by the sum of a constant 300 seconds and a duration drawn from an exponential distribution with a mean of 300 seconds. The impatience of a telephone caller is drawn from an exponential distribution with a mean of 240 seconds. The warm-up for both chat and telephone is drawn from exponential distributions with means of 60 and 45 seconds for telephone and chat, respectively. Lastly, the cool-down durations of chat and telephone are both drawn from an exponential distribution with a mean of 120 seconds.

Moreover, based on current practice at 113, for the experiments, we assume that $n_{\text{triage}} = 5$. Thus, each triage agent can handle a maximum of five triage chats simultaneously. Further, the simulations are trace-driven, and follow the realizations of the time variations of (1) the arrival processes for chat and telephone, (2) the number of triage- and HL-agents, and (3) the fraction of triage call that is forwarded to the HL system.

Fig. 11 shows both the simulated and realized average waiting times for telephone calls as a function of the cumulative number of call arrivals. The results show that for a small number of calls, the average waiting time is rather sensitive to outliers but quickly stabilizes over time. The average waiting time of the data and simulation both converge to the same value, around 80 seconds. This validation experiment is repeated for chat call arrivals, which show similar convergence, see Fig. 12.

In summary, these validation results show that the model works well in predicting waiting times.

V. Demand forecasting

Demand forecasting concerns itself with predicting the call volumes for both telephone and chat calls. The time window

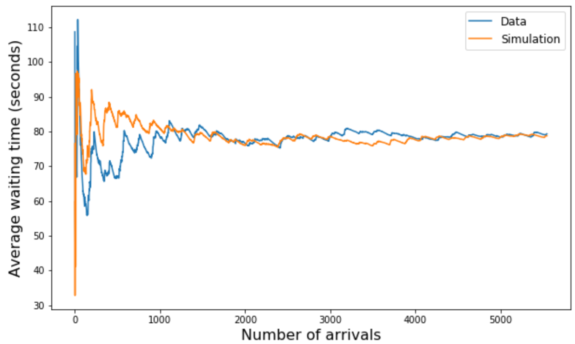

Fig. 11. The average waiting time of telephone calls in the simulation and the data.

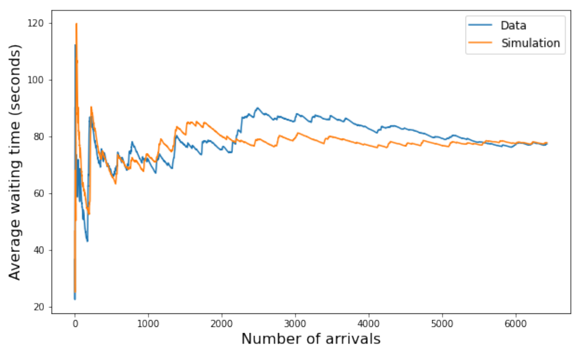

Fig. 12. The average waiting time of chat calls in the simulation and the data.

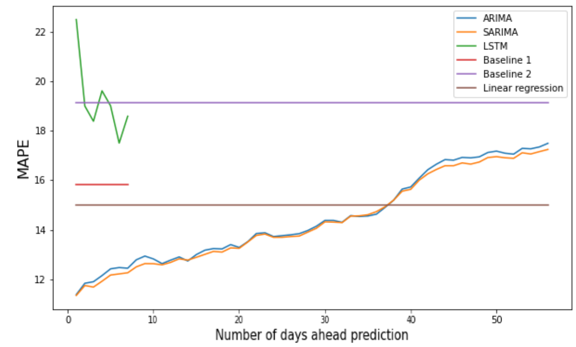

Fig. 13. The error when forecasting telephone.

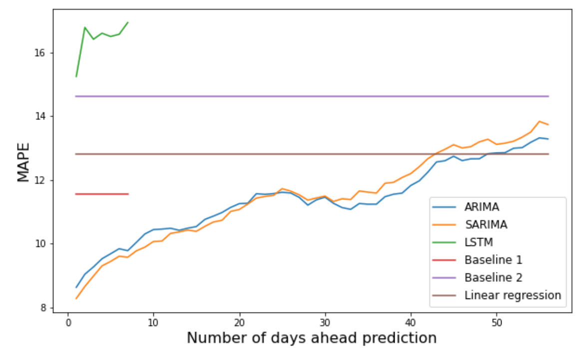

Fig. 14. The error when forecasting chat.

for predictions ranges from 1-day ahead to 8-weeks ahead. Various models and features were considered. More specifically, the following forecasting models were considered: (1) Long Short-Term Memory (LSTM), (S)ARIMA [25], Linear regression [26] and different baseline models. The parameters of (S)ARIMA models are chosen using auto-ARIMA [27], which are $(5, 1, 1)$ for ARIMA and $(1, 1, 1)(0, 1, [1, 2], 7)$ for SARIMA. As baseline, the prediction of day i is the volume measured on day $i-7$ (baseline model 1) and $i-58$ (baseline model 2). The following aspects were seen as important: *trend*, *seasonality* and the *influence of media events concerning suicide*. The provided data shows that there is an increasing trend present and that weekly cycles seem to be predominant. The assumption was that media events also influence the number of arrivals and should, therefore, be included in forecasting. However, our data analysis showed that in most cases, this assumption is invalid (see also Remark 2 above), although some events did have a large effect. However, for forecasting purposes, it is not possible to predict when such an event would take place and if such an event would lead to a long-lasting effect. Therefore, it is chosen to focus on forecasting using historical data of the call volumes. The models will be evaluated using the Mean Absolute Percentage Error (MAPE).

The results of the forecast can be seen in Figs. 13 and 14. We find that (S)ARIMA models perform best (in terms of the MAPE), especially when forecasting for five weeks (or six in the case of chats) ahead or less. For longer forecasting windows, it turns out that a simple linear regression model might provide more accurate forecasts in the case of telephone arrivals. However, both (S)ARIMA and linear regression models have a lower MAPE than the baseline models. The LSTM model has the highest error term in this situation. It could be the case that with more time and optimization, the LSTM will perform better. However, it is questionable if the model would perform better than the (S)ARIMA models. The performance of (S)ARIMA models can be attributed to the flexibility of the models. If an event results in a long-lasting effect on arrivals, the (S)ARIMA model can quickly adapt and include this increase or decrease in demand. An example of how these predictions compare to the actual data can be seen in Fig. 15.

VI. Conclusion & Discussion

This paper proposes a new call center model for mental helplines. The model is validated based on data from [1], the suicide prevention helpline in the Netherlands. The results show that the model accurately predicts the waiting-time performance of the call center. We emphasize that the model is also applicable to other mental helplines with triage and complex conversations requiring warm-up and cool-down periods, possibly with minor modifications.

The warm-up and cool-down durations are estimated mostly based on experience from agents. There is a difficulty in accurately measuring these durations. For future research,

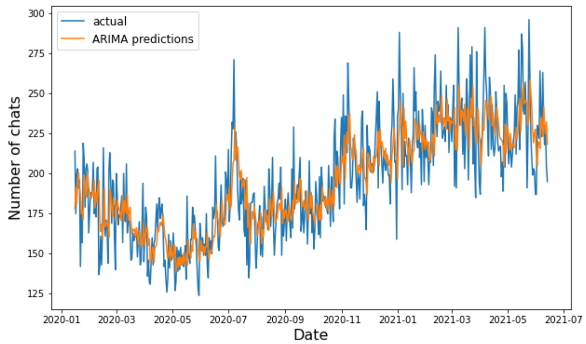

Fig. 15. Comparison between the actual number of chats.

these durations could be based on data and possibly correlated to the duration of the call and wrap-up. This could be done by measuring when agents log off and log back on.

The second contribution of this research is centered around demand forecasting. Various models were tested for forecasting the number of arrivals per day. (S)ARIMA models were among the best in terms of MAPE. We also considered the inclusion of significant media events, such as the suicide of a celebrity. Most remarkably, data analysis shows that these do not have a large or long-lasting effect in most cases. These events cannot be predicted ahead of time. Since the (S)ARIMA models look back at the previous data points, this model is able to adapt if such an event has a long-lasting effect.

This paper aims to provide a complete view, but some aspects require follow-up research. For example, the level of experience of volunteers and paid professionals might affect call durations. Preliminary data analysis into this topic shows no, or at best limited, correlation, but further investigation is needed.

ACKNOWLEDGMENT

The authors would like to thank Robin Costers, Sophie de Vries Robles, Menno Zwart, and Kim Setowski from 113 Suicide Prevention for their great support during this project.

REFERENCES

[1] "Over ons—113 zelfmoordpreventie." [Online]. Available: https://www.113.nl/over-113/over-ons
[2] "De luisterlijn." [Online]. Available: https://www.deluisterlijn.nl/
[3] "Kindertelefoon." [Online]. Available: https://www.kindertelefoon.nl/
[4] J. Scerri, A. Sammut, S. Cilia Vincenti, P. Grech, M. Galea, C. Scerri, D. Calleja Bitar, and S. Dimech Sant, "Reaching out for help: calls to a mental health helpline prior to and during the covid-19 pandemic," International Journal of Environmental Research and Public Health, vol. 18, no. 9, p. 4505, 2021.
[5] M. Brülhart, V. Klotzb üucher, R. Lalive, and S.K. Reich, "Mental health concerns during the covid-19 pandemic as revealed by helpline calls," Nature, vol. 600, no. 7887, pp. 121–126, 2021.
[6] J. Hoogenboezem and T. Traag, "Zelfdoding in nederland: Een overzicht vanaf 1950," Aug 2021. [Online]. Available: https://www.cbs.nl/nl-nl/longread/statistische-trends/2021/zelfdoding-in-nederland-een-overzicht-vanaf-1950?onepage=true
[7] WHO, "Suicide." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/suicide
[8] M. S. Gould, J. Kalafat, J. L. HarrisMunfakh, and M. Kleinman, "An evaluation of crisis hotline outcomes. part 2: Suicidal callers," Suicide and Life Threatening Behavior, vol. 37, no. 3, pp. 338–352, 2007.
[9] J. K. Mokkenstorm, M. Eikelenboom, A. Huisman, J. Wiebenga, R. Gilissen, A. J. Kerkhof, and J.H. Smit, "Evaluation of the 113online suicide prevention crisis chat service: outcomes, helper behaviors and comparison to telephone hotlines," Suicide and Life-Threatening Behavior, vol. 47, no. 3, pp. 282–296, 2017.
[10] N. Izady and D. Worthington, "Setting staffing requirements for time dependent queueing networks: The case of accident and emergency departments," European Journal of Operational Research, vol. 219, no. 3, pp. 531–540, 2012.
[11] M.E. Pratt, "The future of volunteers in crisis hotline work," Ph.D. dissertation, University of Pittsburgh, 2013.
[12] R.C.W.J. Willems, C.H.C. Drossaert, P. Vuijk, and E.T. Bohlmeijer,"Mental wellbeing in crisis line volunteers: understanding emotional impact of the work, challenges and resources. a qualitative study," International Journal of Qualitative Studies on Health and Well-being, vol. 16, no. 1, p. 1986920, 2021.
[13] S. Salmi, S. Méerelle, R. Gilissen, R.D. van der Mei, and S. Bhulai, "Detecting changes in help seeker conversations on a suicide prevention helpline during the covid- 19 pandemic: in-depth analysis using encoder representations from transformers," BMC public health, vol. 22, no. 1, pp. 1–10, 2022.
[14] F. Sundram, T. Corattur, C. Dong, and K. Zhong, "Motivations, expectations and experiences in being a mental health helplines volunteer," International journal of environmental research and public health, vol. 15, no. 10,p. 2123, 2018.
[15] M. D. Christian, "Triage," Critical care clinics, vol. 35, no. 4, pp. 575–589, 2019.
[16] N. Gans, G. Koole, and A. Mandelbaum, "Telephone call centers: Tutorial, review, and research prospects," Manufacturing & Service Operations Management, vol. 5, no. 2, pp. 79–141, 2003.
[17] T.R. Robbins, D.J. Medeiros, and T.P. Harrison, "Does the erlang c model fit in real call centers?" in Proceedings of the 2010 Winter Simulation Conference. IEEE, 2010, pp. 2853–2864.
[18] O. Garnett, A. Mandelbaum, and M. Reiman, "Designing a call center with impatient customers," Manufacturing & Service Operations Management, vol. 4, no. 3, pp. 208–227, 2002.
[19] M. van Buuren, G.J. Kommer, R.D. van der Mei, and S. Bhulai, "Ems call center models with and without function differentiation: A comparison," Operations Research for Health Care, vol. 12, pp. 16–28, 2017.
[20] M. van Buuren, G.J. Kommer, R.D. van der Mei, and S. Bhulai, "A simulation model for emergency medical services call centers," in 2015 winter simulation conference (WSC). IEEE, 2015, pp. 844–855.
[21] M.C.A. van der Burgt, S. Mérelle, A.T.F. Beekman, and R. Gilissen, The Impact of COVID-19 on the Suicide Prevention Helpline in the Netherlands. Crisis. 2022.
[22] A. Weatherall, S. Danby, K. Osvaldsson, J. Cromdal, and M. Emmison, "Pranking in children's helpline calls," Australian Journal of Linguistics, vol. 36, no. 2, pp. 224–238, 2016.
[23] R. Whitley, D.S. Fink, J. Santaella-Tenorio, and K.M. Keyes, "Suicide mortality in Canada after the death of Robin Williams, in the context of high-fidelity to suicide reporting guidelines in the Canadian media," The Canadian Journal of Psychiatry, vol. 64, no. 11, pp. 805–812, 2019.
[24] P. Verboon and E. Schulte Nordholt, "Simulation experiments for hot deck imputation," Statistical Data Editing, Methods and Techniques, vol. 2, pp. 22–29, 1997.
[25] S. Siami-Namini, N. Tavakoli, and A.S. Namin, "A comparison of ARIMA and LSTM in forecasting time series," in 2018 17th IEEE international conference on machine learning and applications (ICMLA). IEEE, 2018, pp. 1394–1401.
[26] B. M. Pavlyshenko, "Machine-learning models for sales time series forecasting," Data, vol. 4, no. 1, p. 15, 2019.
[27] R.J. Hyndman and Y. Khandakar, "Automatic time series forecasting: the forecast package for R," Journal of statistical software, vol. 27, pp. 1–22, 2008.