



Attribution ShareAlike (CC BY-SA) 

## Data Science Project Scoping Worksheet<sup>1</sup>

**1. Project Name:** Detecting Suicide Risk in Chat Applications<sup>2</sup>

**2. Organization Name:** Fundación Todo Mejora

**3. Problem Description:**

### 3.1 What is the problem you are facing?

The Todo Mejora Foundation is a non-governmental organization whose purpose is to promote the wellbeing of children and adolescents who are suffering bullying and suicidal behavior as a result of being discriminated against based on their sexual orientation, identity and gender expression. In Chile, suicide is the leading cause of death among people aged 15 to 24.<sup>3</sup> According to 2015 data, a minor commits suicide every 2.8 days in Chile.<sup>4</sup>

Todo Mejora has a remote assistance service called “Hora Segura” (Safe Hour), which started through Facebook’s messaging service and has the purpose of providing support, guidance, and in specific cases, a referral. A Google grant was obtained in 2015 to have an ad on the search engine every time somebody wrote the words “ways to kill myself”, “painless suicide”, etc. This campaign exponentially increased the demand for the “Hora Segura” service, from an average of 5 conversations a week to 5 daily conversations in 2016. In 2017, a partnership was achieved with Let’s Talk to launch a messaging application and create a platform that allows managing the assistance regardless of whether the conversation requests come from the app, the Foundation’s Facebook chat or the chat on its webpage. The assistance is currently provided at certain times: Monday to Friday from 1.00 to 2.00 pm and 6.00 pm to Midnight; and Sunday from 1.00 to 2.00 pm

---

<sup>1</sup> Created by the Center for Data Science and Public Policy at the University of Chicago and GobLab at Universidad Adolfo Ibáñez. This project description was created by GobLab at Universidad Adolfo Ibáñez as part of a curriculum that is available [here](#). You may use it quoting its creators and adapt it following the Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) License. The terms and conditions are available [here](#).

<sup>2</sup> This worksheet was created by GobLab UAI for teaching purposes. It is a retrospective exercise conducted using public information that doesn’t necessarily represent the project’s initial formulation.

<sup>3</sup> Ministerio de Salud Chile (2019), Programa Nacional de Prevención del Suicidio

<sup>4</sup> Ministerio de Salud de Chile, Departamento de Estadísticas e Información de Salud



and 7.00 to 11:00 pm. Additionally, there are people who try to contact the volunteer team outside of this “Safe Hour”, requests that will be addressed in the next shift. The team is composed of professionals from social sciences, health and education, who voluntarily donate an hour a week to provide this care. The current problem is that the number of users demanding help surpasses the volunteers available for this task at times, so they can’t prioritize contact with children and adolescents at a high risk of suicide. Todo Mejora has even identified external factors that affect the demand for the help service, for instance, the visibility of the service in the context of the Netflix series “13 Reasons Why”, about a case of bullying and harassment that finally ends in suicide. Specifically, Todo Mejora has a team of 60 to 80 professionals who must respond between 8,000 and 9,000 conversation requests a year. Approximately 60% of the consultations received annually are from teenagers in the LGBTIQ+ community, and 40% are from heterosexual cisgender adolescents. 68% of the people who resort to “Hora Segura” have displayed suicidal behavior in the past two months.

**3.2 Who/what is affected by this problem?** (Certain type of people, organizations, neighborhoods, environment)

Children and adolescents (NNA, niños, niñas y adolescentes) seeking support and guidance. Family and close environment of such NNAs.

**3.3 How many people/organizations/places/etc. are affected and how much?** (e.g. mean wait time for surgery, number of students dropping out of school, cost due to tax fraud, etc.)

Considering the age group between 12 and 18 as NNAs, in Chile there are 1,677,106 children and adolescents according to the 2017 Census. Chile, along with South Korea, are the only countries where adolescent suicides have increased instead of decreasing. 1,878 people committed suicide in Chile in 2017, from which 93 were NNAs aged 10 to 19. This number increases to 196<sup>5</sup> between the ages of 20 and 24. 60% of the users of Hora Segura are 19 years old or younger, 24% are between 20 and 24 years old, 9% are aged 25 to 29, and 8% are older than 30.

---

<sup>5</sup> Revista el Sábado. “El silencio de los suicidios en Chile” (2019)



8,714 consultations were received between August 2017 and July 2018, from which 6,902 materialized and among which 2,428 different users can be identified.

### 3.4 Why is solving this problem a priority for your organization?

Because the mission of Todo Mejora is to promote the wellbeing of children and adolescents suffering bullying and suicidal behavior. The main focus is on NNAs suffering discrimination because of their sexual orientation, identity and/or gender expression.

### 4. Goals (in order of priority)

- The technical solution (e.g. predictive model) is not the goal.
- The goal must be measurable.
- Achieving the goal should help solve the problem.
- Typical goals include improving/maximizing/increasing or decreasing/mitigating/reducing some outcome or metric.
- Typical constraints include budget, lack of human capital, legal restrictions, political will and social license.
- Consider tradeoffs between conflicting goals.

	Goal	Constraints
1	To reduce the number of high-risk cases not taken due to the volunteers' lack of time	There is a limited number of care volunteers
2	To reduce the number of high suicide risk cases not taken because they are consultations made outside hours	The program only has 40 hours of care a week

### 5. Actions

- Actions are what institutions can do to address a problem by allocating resources, for instance, inspect facilities, provide preventive services, outreach, etc.



- Ideal actions should help you achieve the goal defined above.

	Action 1	Action 2	Action 3
<b>Action:</b> <i>e.g. inform the owner of a vehicle on how to renew their permit</i>	Establish communication with NNAs who access the platform during the "Safe Hour"	Establish communication with NNAs who need help outside the "Safe Hour"	
<b>Who is executing the action?</b> <i>e.g. IT department (sends email)</i>	Team of volunteers	Team of volunteers	
<b>Who/what is the action being taken on?</b> <i>e.g. vehicle owner</i>	Users of the app, the Facebook chat or the Todo Mejora webpage chat who talk within established hours	Users of the app, the Facebook chat or the Todo Mejora webpage chat who talk outside established hours	
<b>How often is the decision to take this action made?</b> <i>e.g. annually</i>	<i>During the established hours of "Hora Segura"</i>	When a new shift begins after an interruption period	
<b>What channels are/can be used to take this action?</b> <i>e.g. email</i>	<i>"Hora Segura" app, Facebook chat, web chat</i>	<i>"Hora Segura" app, Facebook chat, web chat</i>	
<b>Other useful information about the action</b>		Volunteers don't always have time to immediately address the support requests when a new shift begins, and reestablishing contact is not always possible	



## 6. Data

- The data has to connect to the actions it informs so the organization can achieve its goal.
- Typical data science projects use administrative data as the primary data source and enhance it with publicly available data sources (Census, other open data). Partnering with the private sector or non-profits could be a way to obtain data you might be missing internally.

### A. What data sources do you have internally?

(add columns for more sources if applicable)

	<b>Message records between NNAs and psychologists through the app</b>	<b>Conversation segment labels</b>	<b>Data Source 3</b>
<b>What does it contain?</b> e.g. hospital admission and discharge records at a national level	Historical messages between the user and the volunteer	Risk rating of conversation segments	
<b>What level of granularity?</b> e.g. transaction, person, organization, location	Messages received and sent by the users each time they use the platform	Conversation segment	
<b>How frequently is it collected/updated once it is captured?</b> e.g. in real time, daily, weekly, monthly, yearly, one-off	In real time	Once, for the construction of the algorithm	



<b>Does it have reliable and unique identifiers that can be linked to other data sources?</b> e.g. national identifier	No, it only includes the email of those who make contact through the app	Not applicable	
<b>Who is the internal owner of the data?</b> e.g. hospitals	Todo Mejora	Todo Mejora	
<b>How is it stored?</b> e.g. in a database, PDF, Excel	Database	Database	
<b>Additional comments</b>	Emails may not be real	We need a group of trained volunteers to rate the risk level of the conversation segments	

**B. What data can you get from external, private or public sources?**

<b>What does it contain?</b>			
<b>What level of granularity?</b>			
<b>How frequently is it collected/updated once it is captured?</b>			
<b>Does it have reliable and unique identifiers that can be linked to other data sources?</b>			



<b>Who is the internal owner of the data?</b>			
<b>How is it stored?</b>			
<b>Additional comments</b>			

**C. In an ideal world, is there additional data you would want to obtain/gather that would be relevant to this problem?** (Surveys, CCTV, phone records, DNA, different frequency or granularity for currently available data, etc.)

It would be interesting to have information from the Ministry of Health about the medical and psychological care of the children and adolescents to supplement the care database. However, there is a limitation: not many NNAs seek traditional help out of fear of further rejection.

Additionally, if we could cross-reference comments on Facebook/Twitter and Instagram photos to find other risk factors.

## 7. Analysis

- Typical data science projects include a combination of analyses.
- The analysis is not the goal of the project.
- Choose the right analysis for the right problem.
- You must validate the analysis, and the validation process must match your goal.

	<b>Analysis 1:</b>	<b>Analysis 2:</b>	<b>Analysis 3:</b>
Type of analysis (e.g. description, prediction, detection, behavior change)	Prediction		



Purpose of the analysis (e.g. understand historical behavior of individuals, estimate a patient's risk of disease)	To predict the risk of suicide of an NNA who accesses communication with volunteers through the app		
Which action will this analysis inform?	Actions 1 and 2		
How will you validate this analysis using existing data? (e.g. using historical data, running an RCT)	Training of a predictive model based on historical data. 1,794 conversation segments (pairs), from which 612 were assessed as high risk of suicide. 1,079 segments will be used as a training set, and the validation and testing sets will contain 358 segments each.		

## 8. Ethical Considerations

<p><b>Privacy</b> Are you working with personal and/or sensitive data that is individually identifiable?</p>	<p>Yes, we can deduce the identity of a person through their email. The content of the conversations can reveal personal data (e.g. address) and sensitive data (mental health and others).</p>
--	---





<p><b>Transparency</b> Which stakeholders should know about which parts of the project? (Stakeholders typically include policymakers, frontline workers, people who will be affected by the actions, etc.)</p>	<p>Social Director of Todo Mejora, about the use of the data in the algorithm, and the person who coordinates the volunteers of the Hora Segura program to evaluate the applicability of the algorithm.</p>
<p><b>Discrimination/Equity</b> Are there any specific groups for whom you want to ensure equity of outcomes?</p>	<p>Yes, LGTBI+ children and adolescents have a higher risk than others and they are a priority group for the foundation.</p>
<p><b>Social License</b> If the entire population of the country finds out about your project, will they be ok with it?</p>	<p>It can be complicated, as we would be “prioritizing” care based on an algorithm. We must be very careful with false negatives who may be left without care.</p> <p>We could add a second indicator based on wait time to ensure care for everyone.</p>
<p><b>Accountability</b> Who are the people responsible for all the things above?</p>	<p>Todo Mejora, as the owner of the data (conversations)</p>
<p><b>Other considerations such as consent, legal, etc.</b></p>	



**9. What field trial or randomized controlled trial can you design to validate the project in the field?** The outcomes you will measure should match your goals.

Once the prediction algorithm is designed, different trials can be conducted to measure the efficiency of the instrument: for example, ask the volunteers if they found that the level of risk detected by the algorithm matched the level of risk identified by them.

**10. Who are the external organizations and internal departments that will need to be involved?**

(Typically, data science projects need involvement from data owners, IT infrastructure owners, the problem owner, analytics people)

This worksheet was originally developed by the Center for Data Science and Public Policy at the University of Chicago. For more information about our programs and work, please visit <http://datasciencepublicpolicy.org> or email us at [info@datascienceforsocialgood.org](mailto:info@datascienceforsocialgood.org)

This version of the worksheet has been extended through a collaboration between GobLab UAI, Carnegie Mellon University and ITAM.

GobLab UAI is the innovation lab of the School of Government at Adolfo Ibáñez University. Its mission is to promote the use of data science in the public sector in order to improve public management and have more evidence-based public policies. It trains public servants and does applied research and projects in partnership with government agencies. For more information, visit <https://goblab-uai/> or email [goblab@uai.cl](mailto:goblab@uai.cl)

