



Attribution ShareAlike (CC BY-SA)

Data Science Project Scoping Worksheet¹

1. Project Name: DART. Artificial Intelligence for Blindness Prevention²

2. Organization Name: Teledx

3. Problem Description:

3.1 What is the problem you are facing?

Diabetes is a chronic disease that occurs when the pancreas is incapable of producing enough insulin or the body doesn't use all the insulin it produces. There are two types of diabetes. Type 1 diabetes is caused by an attack of the body's own immune system. It doesn't have a cure and cannot be prevented. Type 2 diabetes is strongly linked to obesity and physical inactivity. The pancreas continues to produce insulin, but in insufficient amounts, and in some cases, it can be cured. Both types of diabetes may cause blindness, renal failure, CVA and even the amputation of legs.

Diabetic retinopathy (DR) is the most common eye disease among those who have diabetes in the world. DR diagnosis requires a test called ocular fundus, in which a medical technician obtains a detailed image of the eyeball that is then analyzed by a specialist.

According to estimations made by Hojman³, in Chile there is a deficit in annual ophthalmologist hours of 39,168, so the necessary human capital to analyze all the ocular fundus exams doesn't exist.

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

¹ 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](#).

² 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.

³ Hojman (2014). El mercado de especialidades médicas de anestesiología y oftalmología en Chile.





People who have diabetes because they cannot have an ocular fundus exam once a year, as per international recommendations. The Ministry of Health, which must spend additional resources in diabetic retinopathy because it is unable to undertake adequate prevention. Ophthalmologists, who have a high workload.

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.)

In 2014, there were 422 million people with diabetes in the world, and the World Health Organization estimates that this number will increase to 600 million by 2040. In Chile, it is estimated that 1 out of 10 people has diabetes. It is estimated that 30% of diabetic patients have diabetic retinopathy.

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

Teledx was created with the sole purpose of technologically supporting the health system's ophthalmological task in order to reduce vision loss and blindness rates in the adult population. The main focus is on diabetic retinopathy, which is why it is imperative to increase diagnosis coverage.

The National Health Strategy developed by the Ministry of Health for the 2011-2020 period sets various goals. One of them is to increase the proportion of people with controlled diabetes. The specific strategy for achieving this goal is to improve the control of diabetic patients through better coverage for various clinical exams, among them the ocular fundus exam⁴.

⁴ Ministerio de Salud (2010). Estrategia Nacional de Salud para el cumplimiento de los objetivos sanitarios de la década 2011-2020.





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 increase the number of annually conducted ocular fundus exams	There is a limited number of ophthalmologists who can analyze ocular fundus exams
2	To reduce the wait time between the exam and the patient's diagnosis	There is a limited number of ophthalmologists who can diagnose diabetic retinopathy





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
Action: <i>E.g. inform the owner of a vehicle on how to renew their permit</i>	Ocular fundus examination	DR diagnosis by an ophthalmologist	
Who is executing the action? <i>e.g. IT department (sends email)</i>	Medical technicians at healthcare centers	Ophthalmologist	
Who/what is the action being taken on? <i>e.g. vehicle owner</i>	People diagnosed with diabetes	People diagnosed with diabetes	
How often is the decision to take this action made? <i>e.g. annually</i>	Annually per person	After the examination	





What channel are/can be used to take this action? e.g. email	Healthcare centers that have the necessary supplies	Healthcare centers	
Other useful information about the action			

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)

	Data Source 1	Data Source 2	Data Source 3
What does it contain? e.g. hospital admission and discharge records at a national level			





Gob_Lab UAI

UNIVERSIDAD ADOLFO IBÁÑEZ



What level of granularity? e.g. transaction, person, organization, location			
How frequently is it collected/updated once it is captured? e.g. in real time, daily, weekly, monthly, yearly, one-off			
Does it have reliable and unique identifiers that can be linked to other data sources? e.g. national identifier			
Who is the internal owner of the data? e.g. hospitals			
How is it stored? e.g. in a database, PDF, Excel			
Additional comments			





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

	Ocular Fundus Exam	Patient Information	
What does it contain?	Ocular fundus exam images	Medical record of the patient (sex, age, diagnosis date, etc.)	
What level of granularity?	Individual level	Individual level	
How frequently is it collected/updated once it is captured?	In real time	One-off at the time of the exam	
Does it have reliable and unique identifiers that can be linked to other data sources?	Yes, name and ID	Yes, name and ID	
Who is the internal owner of the data?	Healthcare centers (Ministry of Health)	Healthcare centers (Ministry of Health)	





How is it stored?	Images	Database	
Additional comments			

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.)

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.





	Analysis 1:	Analysis 2:	Analysis 3:
Type of analysis (description, prediction, detection, behavior change)	Detection		
Purpose of the analysis (e.g. understand historical behavior of individuals, estimate a patient's risk of disease)	To automatically obtain anomalous cases that can turn into diabetic retinopathy		
Which action will this analysis inform?	Action 2. The ophthalmologist will now have filtered information at the time of the patient's diagnosis		
How will you validate this analysis using existing data? (e.g. using historical data, running an RCT)	Training and evaluation with historical data on ocular fundus exams and the subsequent diagnosis		





8. Ethical Considerations

Privacy Are you working with personal and/or sensitive data that is individually identifiable?	Yes, and it is also sensitive data because it refers to people's health condition.
Transparency 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.)	The Minister of Health and heads of healthcare centers must be aware of the project's advances and its subsequent impact. Ophthalmologists must know how to operate the new platform.
Discrimination/Equity Are there any specific groups for whom you want to ensure equity of outcomes?	No, the exams of all diabetic patients will be analyzed. The algorithm could have a different performance for subgroups of the diabetic population, and for that reason it would be good to conduct a disparity analysis.
Social License If the entire population of the country finds out about your project, will they be ok with it?	Yes, because the diagnostic capability for the disease will be increased. In any case, it is important to communicate that the final diagnosis will be in the hands of the ophthalmologists to prevent the resistance of having an automatic system.
Accountability Who are the people responsible for all the things above?	Teledx General Manager Head of the health service where it is implemented





Gob_Lab UAI
UNIVERSIDAD ADOLFO IBÁÑEZ



Other considerations such as consent, legal, etc.	
--	--





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.

The model is trained with historical diagnosis data, but it is important to have a pilot program in at least 1 municipality in order to identify potential risks in the tool's implementation, both in its prediction power and its use by medical technicians and ophthalmologists.

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)

Organization/Department	Description of desired involvement	Name/role of counterpart
Teledx	In charge of creating the image analysis tool	Data analysts
Ministry of Health	Implementation of the tool at healthcare centers	Head of health services networks
Teledx/ Ministry of Health	Training of ophthalmologists and medical technicians on the use of the tool	Training team





Gob_Lab UAI

UNIVERSIDAD ADOLFO IBÁÑEZ



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

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

