THE CURVE

THE DATA CURVE

Smarter decisions through data

M&CSAATCHI

BILL & MELINDA GATES foundation

HARVARD SCHOOL OF PUBLIC HEALTH

dimagi

CONTENTS PAGE

INTRODUCTION

STAGE 1: CONSISTENT DATA USE

STAGE 2: SIMPLE ANALYTICS

STAGE 3: OPTIMIZATION

Technology proliferation is generating large quantities of data

The spread of technology in low-income settings has led to a substantial increase in the digitization of systems across all sectors, including healthcare, agriculture, education, and sanitation. The increase in uptake of technology and digitization is generating a substantial amount of data across all of these sectors.

In recent years, researchers and practitioners have begun to shift their focus from digitizing interventions to extracting value and insights from the near-real time data that is being collected. A range of techniques, from monitoring and plotting incoming data to outlier detection and cohort analysis are being deployed to derive value from these datasets. The majority of this value comes from consistent data use and simple analytics. In specific cases, however, machine learning is being used to optimize specific problems and extract a final few percent. Problems such as computer vision for agricultural development to automatically determine pests, digital signal processing to detect signatures in audio data in the health sector to automatically classify coughs as asthma, bronchitis, or pneumonia, or natural language processing to triage health messaging systems.

Fig 1

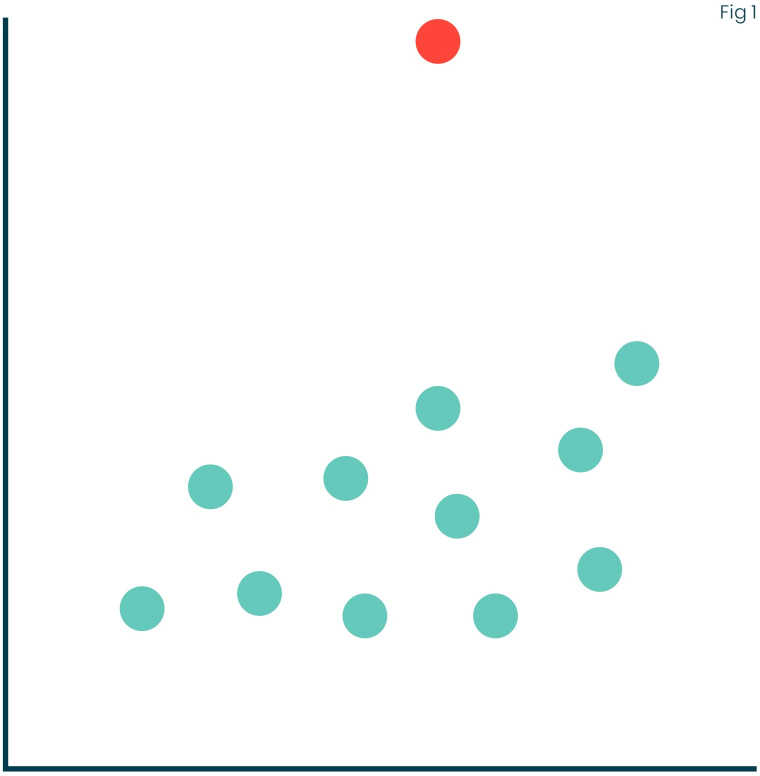


Figure 1: Outliers in the data can be a valuable signal that something unexpected is occurring programmatically.

Better use of data will enable organizations to make better decisions

To guide organizations towards making better use of their data, we propose the data curve: a guide to increasing the value derived through consistent data use and appropriate analytics. It is essential that groups develop strong, consistent data use habits before they begin to use simple analytics for data-driven decision making. Similarly, organizations should fully explore a range of analytics options before attempting to optimize data use using complex machine-based approaches. As described in the third stage, some organizations will find limited value for the respective effort required during the optimize stage.

Practical steps on the Data Curve

This guide is broken down into three distinct stages: consistent data use, simple analytics and optimization (Figure 2). Despite the presentation as three distinct levels, in reality they are three phases, each with their own complexities. There is a more complex continuum of data use skills and challenges underlying the data curve.

With this guide, it is our aim to help organizations get better faster, by showing the value of each stage, as well as the steps required. For each stage of the data curve, we will provide:

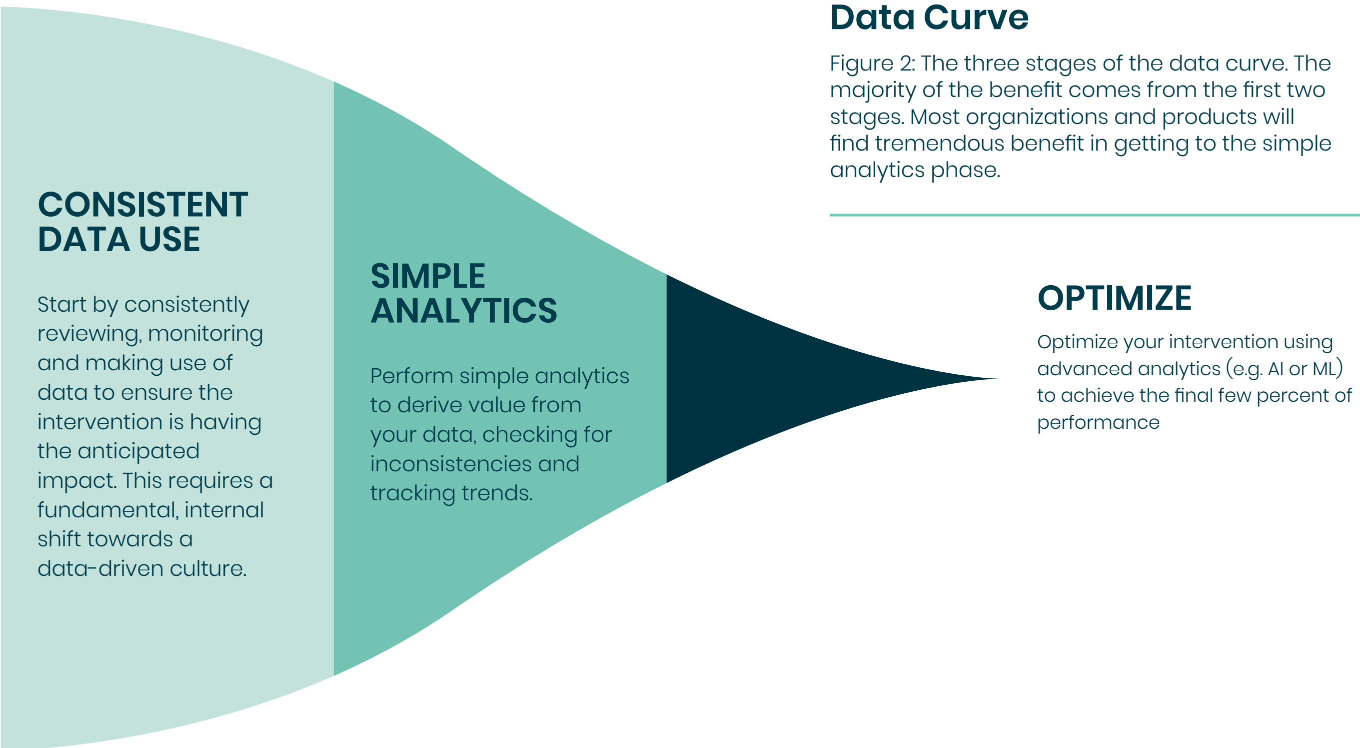
- **Benefits** – Introduces the benefits of this stage.
- **Data use steps** - Explains the stage and gives concrete, but general steps of how to use data at this stage.
- **Example techniques** - Techniques that can be used at this stage of the data curve to draw out actionable insights.
- **Promising Practice** - Case studies and practical examples of organizations using data at this particular stage of the data curve.

Assumptions for this guide

This guide assumes that a digital intervention is generating data, either through explicit collection, or as a byproduct of the intervention. This is a necessary precondition to any of the stages of the data curve. Data collection may occur with frontline workers (FLWs) (e.g., in a health or agriculture setting), or can occur as a part of a larger data tracking system during routine service delivery (e.g., an electronic medical record in a clinic or hospital, or a system to track marks in an education setting).

There is a further distinction between static (e.g., census data or other once-off data collection efforts) and operational datasets (e.g., routine and ongoing data collection). While this guide and the data curve are primarily targeted for use by organizations with operational data collection, many of the principles will apply to a static dataset as well.

Fig 2



Data Curve

Figure 2: The three stages of the data curve. The majority of the benefit comes from the first two stages. Most organizations and products will find tremendous benefit in getting to the simple analytics phase.

Key takeaways from this guide

The goal of this document is to provide organizations with a pathway towards deriving more value from the data they collect. We will provide inspirational case studies of the appropriate application of analytics in the development sector. A key takeaway is that there is tremendous value in the cultural shift towards consistent data use and basic analytics, which push an organization towards data-driven decision making.

STAGE

1

CONSISTENT DATA USE

The first step of an organization's journey along the data curve is to develop consistent data use habits. During this initial step, organizations should regularly be checking to make sure that data being generated is completed and entering the system correctly. The organization should then ensure that data use is occurring - that actions are taken based on the data that was collected. This is the most difficult and time-consuming stage of the data curve. Most organizations will begin at some point along the spectrum included in Stage 1.

Benefits of consistent data use

- **Data aggregation** – If your data is spread across multiple systems, this step will ensure that you have a repeatable way of combing it together in a single location.
- **Data quality** – Initial checks on the data will ensure that one can catch data errors before they become a larger problem. It is common that there is a tight feedback loop between the questionnaire creation, training of FLWs or enumerators, and consistent data use.
- **Internal alignment** – Consistent data use will lead to increased visibility and quantitative awareness of what is happening with the project. This step will ensure that team members are all on the same page. The cultural shift required for consistent data use will mean people know where to look for data and gain understanding about what is happening in the field.

Who is using and tracking data outputs?

Similar to how organizations ensure they define what their sales team would consider a "qualified" lead—the person organization that could eventually purchase a product or service—data use cases should aim to define a "qualified" data user before investing in analytics or AI/ML. A well-qualified data user will have the authority and resources to utilize the data output (e.g., routine report or dashboard) when it is created and refreshed. If a qualified data user cannot be defined for a given data use goal, it means that in addition to creating the data output, a new role must be created—or an existing role expanded—within the organization to make the data actionable and to achieve the desired outcome. In short, it is essential that organizations identify a qualified data user for each data use case.

Data Use: Five steps to consistent data use

Consistent data use can be broken down into five main steps. These are essential to moving a project towards a data-driven approach. It is important to note that these are not meant to be a sequential set of steps, but rather will require iteration and a number of cycles. For example, during regular data review (step 5), additional reports (step 4) may be identified, which require new unified views of data (step 3) to compute.

Step 1: Collate your data

The first step is to identify both internal and external data sources that are relevant to your project. Internal data sources may include any data you are collecting about beneficiaries, data about sites where you are working, and data about those who are collecting the data (e.g., agriculture extension workers or frontline health workers). External data sources may include census and other demographic data, weather data, or data collected by other projects and organizations operating in the same locale as your project. In addition to gathering relevant data, it is important to omit data that is not relevant.

If the data is being pulled from multiple sources, the gathered data should then be loaded into a common data repository. As a start, this could be as simple as a folder that contains a number of spreadsheets, but organizations will see the most benefit from investing in and managing a database to hold all of this information further down the line.

Finally, it is important to identify how different data sources relate to one another. Unique identifiers and common keys (e.g., location names, patient ID numbers) must be identified and matched between different data sources.

Fig 3

Figure 3: There are many approaches to disparate sources of data. For analytics, we have found the data warehouse approach to be most useful where all data is replicated into a shared location for analysis. Alternatively, tools like Superset and Metabase can be used to pull data from external sources to view on a common dashboard, which is a great solution to get started looking at data quickly.

Data from a single source

If the data of interest is all coming from a single source, one can jump straight to the next step. The sanity check and analyses can likely be done "in product," as opposed to the data warehousing approach we are recommending. This will be faster, easier, and more sustainable as it does not involve a separate system to maintain the data.

Step 2: Sanity check on data

Once all of a project's data has been gathered together in a single location, the next step is to dive into the data to gather a high-level, preliminary understanding of the data. At this stage, data managers should focus on individual fields to look for missing data, clean any values that are known to be erroneous, and circle back to adjust the data collection process if necessary.

Step 3: Create unified views of the data

During this step it is necessary to export, transform and load (ETL) data into a new location in order to create unified views. We provide additional details on ETL below. The goal of this step is to build a unified view of the data by combining the different data sources together and aggregating if necessary. For example, it may make sense to create an aggregated, unified view of deliveries for a catchment area, providing one row per delivery with details about the pregnancy and delivery outcomes in the columns.

Step 4: Build reports

Drawing from the unified views of the data, projects will find benefit from creating regular reports to visualize or further aggregate data. A number of example techniques are given below. The goal of this step is to produce outputs that will be reviewed on a regular basis in the subsequent step.

Step 5: Routine data use

The final step is to consistently use the data during routine data review. This data review should occur with all levels of the program team and should explicitly not be limited to just the technical or data team. The frequency of data reviews will be context and project dependent and may change over the course of a project.

Iteration required

One anonymous organization that Dimagi worked with in the past was doing household surveys, including asking about the building material used for the walls in the home. When they began to do simple sanity checks on their data, it was discovered that the distribution of answers did not match expectations. Following up with enumerators uncovered that the available options did not match the reality: there were options for "sticks and mud" as well as "brick", but no option that matched the plastered walls that were common in the area.

Performing ETL 'in product'

As mentioned in step 1, it will be easier, faster, and more sustainable for projects with a single source of data to perform this step "in product." This is, however, dependent upon the product being used and whether report generation can be used to perform the ETL steps mentioned. In some straightforward cases, it may even be possible to import simple external datasets for use in product.

Creating a unified view

The extract, transform, load (ETL) process comes from the database and data analytics community. It describes the high-level steps required for producing a unified view.

- **Extract** – In the first two steps of consistent data use, we are performing the extract step of ETL. That is, we are gathering potentially disparate data sources together and ensuring that the data looks as expected.
- **Transform** – The transform stage will collapse or aggregate data. If we were creating a table of deliveries, for example, one may not need to include every single antenatal care (ANC) visit and every detail of what occurred during those visits, but could perhaps aggregate to keep the total number of ANC visits, or whether certain healthy actions were taken.
- **Load** – The final stage is to load the transformed data into a system that supports querying and aggregating data together. For most groups, we recommend investing in setting up and managing a relational database if this is something that cannot be achieved in product.

Culture shift

Consistent data use demands a dramatic culture shift within a project and organization.

It is necessary that there is alignment at all levels of the project or organization around data use. While the aforementioned qualified data user should be driving data use, it is not sufficient for people on the data and tech teams to be the ones using data.

Through continued review of the data, an organization or project will hit a tipping point where a sufficient number of people are using the data frequently enough—and deriving sufficient value—that it becomes the first place that people will go when looking for an answer to their query. The data may not be perfect in the first iterations, but it will still be useful to track how indicators change. And, more importantly, this will build the habit of routine data use.

There are a number of benefits to this culture shift towards routine data use, including increased data quality. It is easier to catch data errors before they become a larger problem. It is common that there is a tight feedback loop between the questionnaire creation, training of FLWs or enumerators, and consistent data use.

Examples technique

Here we provide an example set of techniques that might be used in the consistent data use section.

This is not meant to be comprehensive, but rather an example of the types of reports that we might expect an organization be generating for use in the routine meetings:

- **GeoMaps** – Graphs that plot data that contains geographical information on maps can be useful to see how things are changing from meeting-to-meeting and to compare simple geographical differences. Agriculture and health data are commonly plotted in this way to understand how different geographical regions are performing and can easily spot discrepancies or missing data.
- **Temporal graphs** – Creating graphs that chart indicators over time can be important for noticing major shifts in performance or sudden drops in indicators that are indicative of other programmatic issues. They are also essential for getting better faster and monitoring
- **Dashboards** – A common request for programmatic monitoring is the creation of dashboards that are an amalgamation of one or more techniques. These often combine multiple reports together in a single view and can be the single source of data for a routine data review.

Fig 4

Figure 4: The COVID-19 dashboard from the Centre for Systems Science and Engineering at Johns Hopkins University is an example of a feature-rich dashboard with a GeoMap in the centre.

Promising practice: Stage 1 in use

The ColaLife project, which works to bring life-saving Oral Rehydration Salts (ORS) to remote locations in low-income settings, has benefitted tremendously from consistent data use. In mid-2014, the ColaLife project found that they were capturing digital data on their intervention after they standardized their frontline data collectors with tablets.[i] However, “analyses took a long time to do and so were infrequent.” The group worked to develop a dashboard of their data within the next year (step 4), but found that despite the energy that went into the dashboard, it wasn’t until they gathered all of their staff together in early 2016 to implement a culture shift towards routine data use that they began to see the benefit. The weekly release of the dashboard is now seen as “an event” within the organizations, with teams eager to see how they are performing.

The ColaLife team reported that viewing performance on the dashboard “had a huge motivational effect, with the different district teams comparing their performance.” District teams that were not performing well would look to learn from the higher-performing colleagues. They specifically took steps to develop and spread “‘route planning’ across all districts” in an effort to increase efficiency in their work.

STAGE
2

SIMPLE ANALYTICS

As routine data use grows and the culture shift towards data-driven decisions sets in, there will be a natural increase in the complexity and specificity of queries. As mentioned in the introduction, while we present these as discrete stages, it is a continuum. The boundary between more advanced reports for routine data use and simple analytics is blurry at best. That said, we do have specific recommendations for how to engage and strengthen simple analytics.

Benefits of simple analytics

- **Interrogation of data** – As we move towards simple analytics we begin to start interrogating the data with more complex or time-consuming queries in order to pull answers from our data.
- **Data-driven decision making** – Simple analytics are required by data-driven decision making. Looking at data leads to more questions, which often require different analytical techniques to answer.
- **Increased confidence** – Interrogating and viewing data through the lens of different analyses will lead to an increased confidence in the impact that the project is creating.

Data Use: Four steps to simple analytics

Below, we outline four steps to using simple analytics on your data. Because of the wide range of analytical techniques available, these steps are necessarily broad. Nonetheless, it is our experience that following these four steps can have a positive impact in simple analyses of an organization's data.

Step 1: Define your goal and metric for success

The first step for simple analytics is to clearly define your goal and the metric you will use to measure success. It is important to not underestimate the importance of, and time required for, this step. Being able to clearly define the specific goal is an important step for both simple analytics and more advanced AI-driven analytics. If the goal is to select an analysis that is intended to be actionable, it Is also Important to define what users will be taking action to achieve the goals.

Step 2: Pick your technique

The second step of the simple analytics stage is to select a technique. When selecting a technique, it is important to match that technique to the goal and metric that were previously selected. Focus should be given to the question: will this technique answer the specific question we are asking? Example techniques are enumerated in the following section.

Step 3: Iteratively apply your technique and investigate data

It has been our experience that almost every investigation leads to more questions. For example, imagine we are exploring a correlation between FLW performance and health outcomes. It would not be unusual in our experience to discover that the answer is that there appears to be a weak correlation. A reasonable next step might be to evaluate whether the correlation changes in different locales, or with different ages of FLWs, or any additional potential confounders that are contextually relevant.

Step 4: Evaluate if you've met your goal

The metric and the goal defined in step 1 will be used to evaluate the success of the simple analytics. Was the original question answered? What additional questions have been raised? What additional analytics should be performed? If the analysis was Intended to be actionable, did it result in the actions being taken?

Examples techniques

There are a wide range of different techniques that fall under the simple analytics stage. Below we provide a sampling of these techniques, stressing that this is not meant to be an exhaustive list.

Correlations

It is common to look for the "relatedness" between different variables in your dataset. This will often be between input and output variables—for example the FLW age and the amount of work s/he is performing. Correlation analysis can be extremely helpful, though it is imperative to keep in mind that correlation does not imply causation. There are a number of ways to explore correlations, including:

- **Statistically** – Pearson and Spearmen are the most common statistical tests for correlation. As with all statistical tests, it is essential to understand and check the necessary preconditions for validity.
- **Crosstabulations** – This is a simple NxM matrix for categorical variables to measure how many rows in your dataset exist for each combination of the two variables. It is possible to make crosstabs for more than two variables, but it is not advised as it can quickly become unreadable.
- **Heat maps** – Another popular method for exploring correlations is to generate correlation numbers between variables and colour-code the numbers according to the level of correlation to generate a heat map. This can be an excellent way to visually see the relationship between a medium number of variables at once. However, if there are too many variables it can be difficult to interpret as the natural instinct to look for higher-level patterns will be dependent upon variable ordering.

Fig 5

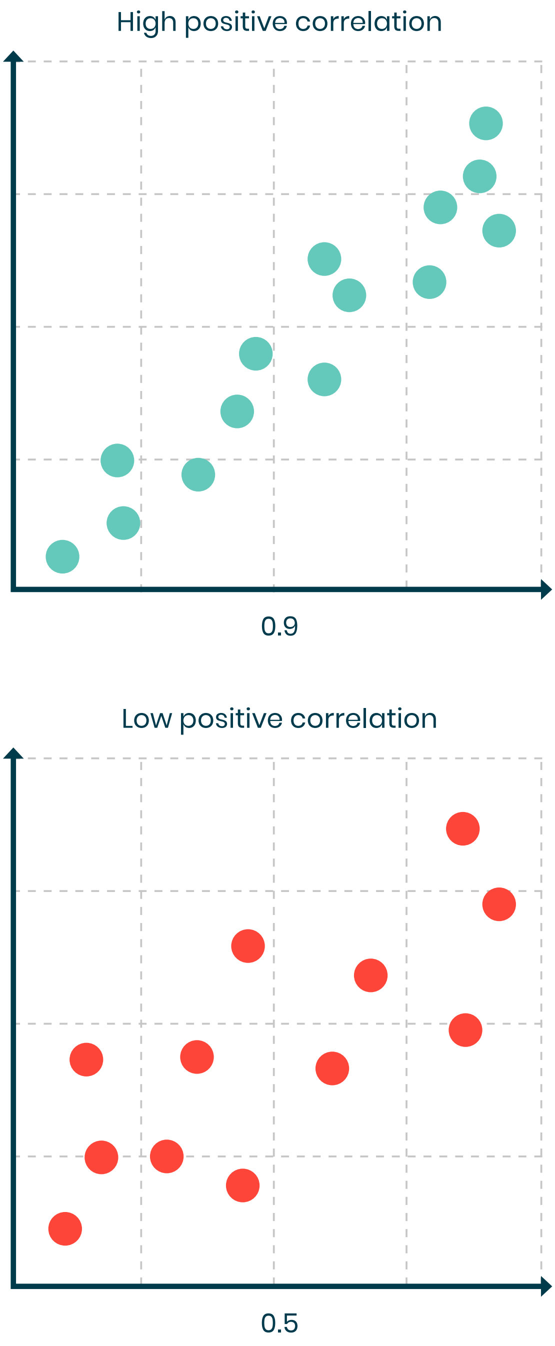


Figure 5: An example of a high-positive correlations and a low-positive correlation. Statistical tests can be one way of quantitatively assessing correlations, but often just plotting the data to check for trends and potential correlations can be the most useful for informing decision making.

Cohort analysis

Splitting your dataset into subgroups for further analysis is a very common technique. Segmentation can be based on groups that receive different interventions, groups that started receiving an intervention at different time, groups that share demographics and so on. There is no limit to the way that groups can be partitioned, though it is important to think about any bias that might exist in partitioning. For example, if all of the highest-risk patients began to receive an intervention first, then there is a potential confound between start date and patient risk. Once appropriate groups have been selected, the metrics for each group can be compared.

Outlier or anomaly detection

Simple techniques, including threshold detection and statistical techniques such as a Chi-squared test, can be used to identify outliers and anomalies with a dataset. Additional investigation is almost always required for anomalies detected in a dataset. It is important to keep in mind that the cause of the anomaly can come from any number of things including typos, training inconsistency, satisficing, and true anomalies. Performing outlier detection can help target additional investigations required.

Promising practice

Dimagi—in collaboration with Catholic Relief Services, Vatsalya, Harvard University and the University of Washington—developed a self-tracking application for frontline health workers in India.[ii] The application allowed frontline workers (FLWs) to see their performance—measured using the simple heuristic of the number of home visits performed—compared to a random subset of their peers. The analysis was available to the FLWs, resetting with the calendar month to start the counts again. Compared to a control group, the intervention—known as ASTA—increased the number of home visits performed by over 20% .

Fig 6

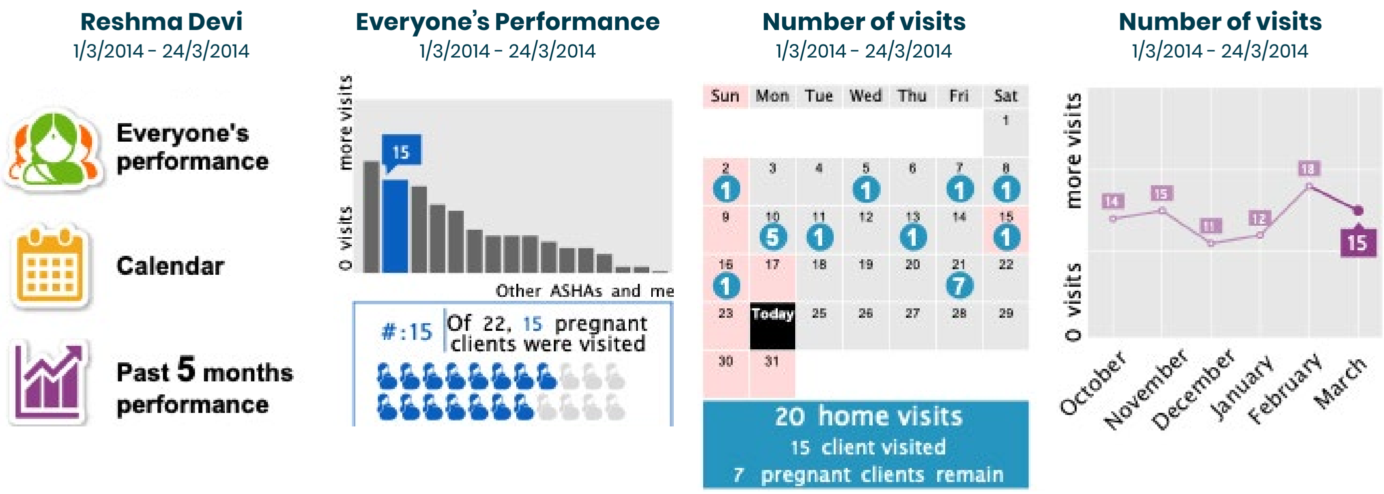


Figure 6: The ASTA intervention self-tracking application that allowed frontline workers to view their performance compared to a random subset of their peers, in addition to daily data and monthly historical performance.

STAGE

3

OPTIMIZATION

The final stage is about optimizing, which for many grousp will mean applying Artificial Intelligence (AI) or Machine Learning (ML) to their projects. AI is broad term encompassing a number of subdisciplines, including natural language processing, natural language generation, and computer vision, focused on using computation to mimic human-like intelligence.

In this guide we focus on supervised machine learning, which can be defined as "data-driven predictions" from labelled data. Building on the work done by others, including the excellent report released by USAID[iii] we believe strongly that the role of AI is to augment human actors. This builds on Kentaro Toyama's theory of amplification[iv], which states that technology in development is an amplifier of human intent.

In an effort to avoid duplication with other available resources, we explicitly consider several topics out of scope, including: how and when to apply artificial intelligence and machine learning[v], as well as an enumeration of the different types of bias in machine learning and how to mitigate them[vi].

Finally, this stage of optimization—as illustrated by the UPS example below—is primarily applicable for organizations and products that need to tune their systems to achieve a final few percent of performance.

Benefits of optimization

- **Increased impact** – Amortized over a large scale, the relatively small gains of optimization can achieve greater performance from systems.
- **Automation** – In many cases, certain aspects of AI—e.g., computer vision or natural language processing—can automate mundane repeatable tasks that previously were done manually. The success of these automations will likely vary based on context and availability of training data.
- **Hidden patterns** – The optimization stage is often able to unlock hidden patterns in the data that were not previously thought about. For example, machine learning can look for the relationship between a very large set of input variables.

When do you need to optimize?

We believe that for most interventions and projects, the consistent data use and simple analytics stages will provide the best effort to value ratio. The final stage, optimizing the intervention, is often only useful for squeezing the final few percentage of value and is most useful when the relatively small improvements are amortized over massive scale.

As an example, United Parcel Service (UPS), a package delivery courier in the United States, made use of advanced analytics and continuous learning to build a heuristic to calculate driver routes[vii]. By combining a number of small optimizations in route planning, including avoiding left turns to decrease accidents and save on fuel and time, and adding those small savings up over millions of deliveries per year, the system saves the company hundreds of millions of dollars per year. The large gains achieved by this optimization are the result of their scale and are only enabled by the fact that the organization had a robust delivery system available. On a per-transaction level, the savings are relatively small.

Similarly, we believe that the application of advanced analytics, AI, and machine learning to problems in global health and similar sectors are most appropriate for existing, robust, scaled interventions that require additional optimization at the cost of significant work.

Fig 7

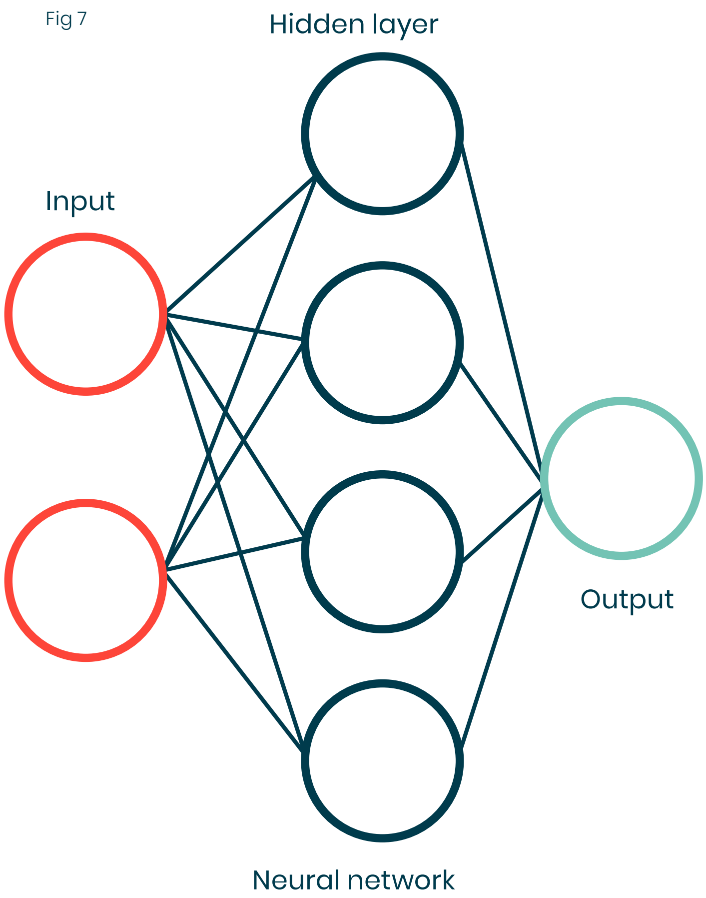


Figure 7: A neural network is one approach for developing machine learning models that has proved to be very successful in image and voice recognition tasks when there is a substantial amount of data. However, the hidden layers mean that explainability is difficult. It is important to fully define your use case and understand what is important when selecting an approach.

Step 1: Define your goal and metric for success

The first step for machine learning parallels that of the simple analytics stage. One must clearly define the goal and the metric that will used to measure that success. In supervised machine learning applications, one must ensure the labels are explicitly and clearly defined as well. For example, if a project was to try to classify urgent versus non-urgent SMS messages, it is essential to have a consistent and standardized labeling process for assigning an urgent or non-urgent label to the messages.

There are a number of standard metrics to evaluate machine learning models, including accuracy, precision, recall, and the area under the ROC curve (AUC). Different goals will need to balance these and other metrics appropriately.

It is also essential to identify and empower the qualified data user for this optimization stage. This is equally, or more, important than identifying the qualified data user during the previous steps due to the advanced nature of the analytics.

Step 2: Get as much data as possible (then gather some more)

In some cases, where it is possible to use prebuilt models, no additional data is required. Doing text-to-speech in English, or basic object recognition of images may be two such use cases. However, when one wants to build a model from scratch, or modify a prebuilt model to make it domain-specific, one must gather new labeled data to feed into the machine learning algorithm.

Step 3: Clean data, think about bias and balance

One core rule in machine learning is that garbage in equals garbage out. This statement covers a large number of potential problems in machine learning, including poorly labelled or cleaned data, as well as bias in your data set. Machine learning is particularly susceptible to carrying any biases in training data forward into the predictive model. For example, imagine training a self-driving car during the day, but expecting it to operate at night as well. In this case, the car should have been trained on both daytime and night-time data.
Finally, it is important to think also about balance in the dataset and how they affect the metrics used for

How much data do you need?

One of the most common questions we hear from groups that want to use AI or machine learning is "How much data to I need?" Unfortunately, this is a highly context-specific question without a clear answer. One way to gather intuition is to think about two different computer vision problems. In the first problem, the goal is to classify an image into two categories that are sufficiently different, and the images within a category sufficiently similar—for example, labeling an image as a "shark" versus a "dog"—and in the second problem, the variation between categories is more limited—for example, labeling images as a "golden retriever" versus a "yellow labrador". In the first example, one could imagine using colour, or environment, or number of legs as a method for segregating images. In the second example, the segregation is more subtle, and therefore more example images are required.

evaluation. For example, if you are trying to classify a rare event, the algorithm could achieve high accuracy by assuming that rare event never occurs. There are a number of different approaches, including under-sampling majority classes or over-sampling minority classes to achieve a more balanced dataset. The precise mechanism will need to be context dependent.

Step 4: Set aside some data for final evaluation

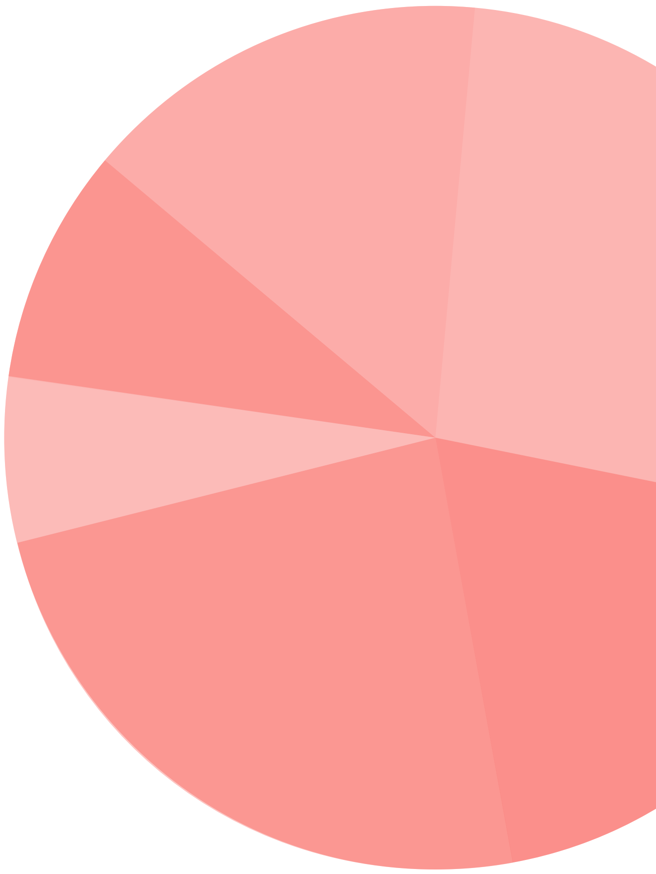
Best practices in machine learning dictate that before building any models, you should segment your dataset and set aside a subset of the data for a final evaluation at the very end of your model building process. Depending on the balance of the dataset, a stratified segmentation of the data may be more appropriate to ensure appropriate balance in both datasets.

Step 5: Iterative on model building and parameter tuning

The next step is to iterate on building models using the remainder of the dataset. This dataset will be further divided into a training and test set, with models being built from the training set and evaluated against the test set. During this iteration process, different algorithms and different parameters for those algorithms can be used to try to build higher-performing models. For many use cases, this process can be automated across a predetermined search space.

Step 6: Evaluate if you've met your goal

Once you have completed the model development and tuned all necessary parameters, best practice is to evaluate that machine learning model against the previously withheld sample of data (step 4), which would not have influenced the model that was built. This use of "unseen" data by your model approximates how it would perform in the real world on new data. If this is an operational dataset where we are continually adding new data, one can use new, unseen data to evaluate.



Promising practice

Case study 1: Using computer vision to identify agricultural issues[viii]

Plantix is a mobile-phone based application that allows farmers to take photos of crop issues, including diseases and pests. A server-side system then uses computer vision to try to classify the specific issue and recommend an appropriate course of action.

Case study 2: Using NLP to triage health messages[ix]

Jacaranda health is using machine learning to triage incoming health messages from new mothers in an effort to identify and surface important and urgent messages as quickly as possible.

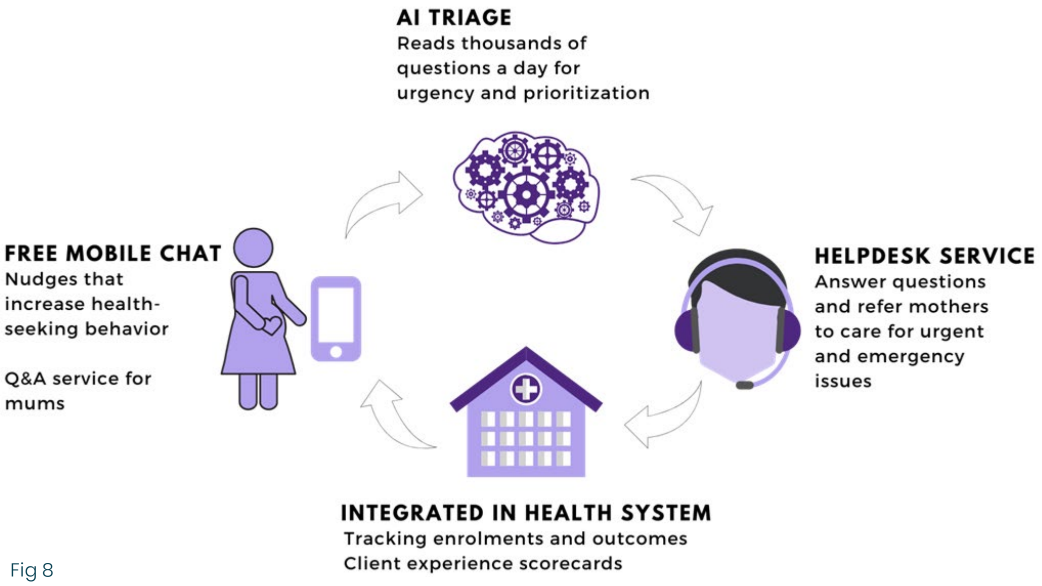


Fig 8

Figure 8: The workflow for how AI is used to optimize messaging systems at Jacaranda Health. In this case they had an established project where they were consistently monitoring and analyzing their program. AI was introduced to optimize performance and improve responsiveness of urgent messages in their call center.

i Barry, S 2016, Dashboards as a motivational tool, colalife, weblog post, 19 September 2016, viewed 19 July 2019. <https://www.colalife.org/2016/09/19/dashboards-as-a-motivational-tool/>.

ii DeRenzi, B., Wacksman, J., Dell, N., Lee, S., Lesh, N., Borriello, G. and Ellner, A., 2016, June. Closing the feedback Loop: A 12-month evaluation of ASTA, a self-tracking application for ASHAs. In Proceedings of the Eighth International Conference on Information and Communication Technologies and Development (p. 22). ACM. DOI: https://doi.org/10.1145/2909609.2909652

iii Rockefeller Foundation and USAID Center for Innovation and Impact (2019) Artificial Intelligence in Global Health: Defining a Collective Path Forward. Washington DC: USAID CII's Innovating for Impact Series. Available at: https://www.usaid.gov/cii/ai-in-global-health (Accessed: 19 July 2019).

iv Toyama, K., 2011, February. Technology as amplifier in international development. In Proceedings of the 2011 iConference (pp. 75-82). ACM. DOI: http://dx.doi.org/10.1145/1940761.1940772

v https://www.gov.uk/guidance/assessing-if-artificial-intelligence-is-the-right-solution

vi Paul, A., Jolley, C. and Anthony, A., 2018. Reflecting the Past, Shaping the Future: Making AI Work for International Development. USAID, Washington.

vii Holland, C., Levis, J., Nuggehalli, R., Santilli, B. and Winters, J., 2017. UPS optimizes delivery routes. Interfaces, 47(1), pp.8-23.

viii Plantix (2019) Best Agriculture App. Available at: https://plantix.net/en/ (Accessed 19 July 2019).

ix Jacaranda Health (2019) SMS Program. Available at: https://www.jacarandahealth.org/prompts-program (Accessed: 19 July 2019).