



MTXT.io Advanced Analytics:

MEMOTEXT | 7 BELMONT ST. TORONTO, ON | 1877.MEMO.TXT | @MEMOTEXT |



problem, the KPIs must be well defined and be measurable from the data.

2. Data Collection and Cleaning

All inputs and drivers that are related to the business problem should be leveraged. After available data sources have been identified, data is extracted in a functional format to be stored in databases. The raw data is then transformed and cleaned in preparation for analysis. Data cleaning can include:

- removing duplicate, inaccurate, or irrelevant data
- fixing structural errors (ex. Mislabelled categories)
- removing outliers and/or suspicious measurements
- handling missing data
- normalizing data (ex. Date-time components)

3. Data Exploration and Model Building

Exploratory analysis is conducted to understand the data set in terms of descriptive statistics and any high-level trends, correlations or other relationships. The trends and patterns seen in this phase can serve as new ideas that may help in solving the business problem or identifying new business opportunities which will often lead to additional KPI definitions. The goal of data exploration is to gain a meaningful understanding of the data and understand any limitations.

Within this data exploration phase, features start to be engineered from the data. A feature can be defined as a measurable property of a given thing, event, or experience that is being observed. It is also more commonly referred to as a field, characteristic, data element, or variable, and eventually also being known as a 'predictor' in a machine learning model.

More specifically, feature engineering (Figure 2) refers to the process of using domain knowledge of the business problem to create derived proxies and extract distinguishing characteristics from the data that will improve the performance of machine learning algorithms (i.e. adherence archetypes, provider complexity, clinical events from medication changes, etc.). The goal with model building is to "model" the real world which is more difficult than it may appear. The world around us has many subtleties and complexities that can be difficult to account for. As such, the feature engineering process is very cyclical and iterative in nature and requires ongoing consultation with the stakeholders and individuals who have the domain expertise (i.e. clinicians, pharmacists, etc.). It is an agile collaboration in which you will determine which features are the most important to capture to use in future modelling. If new ideas arise, you may find yourself back at the drawing board.

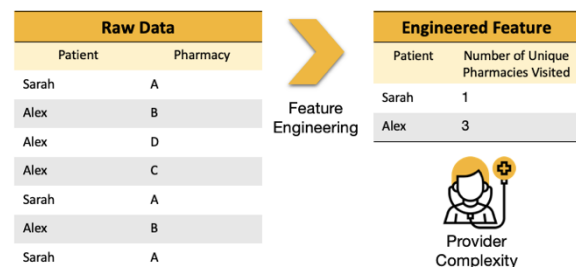


Figure 2. A simplified depiction of feature engineering to create a Provide Complexity feature from raw data.

Moving forward, the most suitable machine learning algorithms as well as the strategy for evaluating the performance of models are selected. Depending on the business problem and KPI definitions, different algorithms may be selected. From a high level, there are two main types of machine learning that are used when working with healthcare data – supervised and unsupervised learning.

Supervised learning refers to data that is labelled, meaning you know its classification or result. The algorithms are trained on cases that



| EVIDENCE | EMPATHY | EXPERIENCE |

have a known outcome (i.e. patient hospitalization) and then those learnings are applied on new, or unseen, data to make a prediction for those cases (i.e. given this set of patient features, will they be hospitalized or not?). Examples of algorithms include logistic regression, linear regression, Naïve Bayes, decision trees, random forest, support vector machine, and gradient boosted machines.

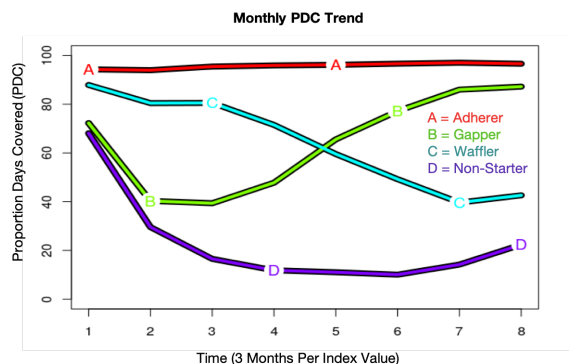


Figure 3. Longitudinal Clustering - PDC using sliding 90-day intervals across a 24-month period with medication carry-over allowed to identify different adherer types (N=12,505).

Conversely, unsupervised learning refers to data that is unlabelled. This type of machine learning is used for finding unknown patterns and hidden structures within the data, without explicitly stating what is being looked for. Examples include K-means clustering, hierarchical clustering, association mining, etc. (Figure 3).

Once the machine learning models have been built, they must be validated to fine-tune the parameters and then tested on new (withheld) data to assess performance.

4. Target Population and Intervention

The results of the data analysis are presented to stakeholders and interim reports are provided along with answers to the original business problem and hypotheses. Steps 3 and 4 are subject to iteration as the models are

adjusted when new findings or questions arise. The final data is summarized in the form of key takeaways and recommendations that can be actioned upon.

Next, it is time to determine which patients to target for the intervention. This requires circling back to the original business problem and completing a root cause analysis. The right patients may be the ones who are at the most risk – whether it be the risk of hospitalization, becoming high cost, developing a complication, or switching off of a certain therapy. Deploying an intervention to the entire patient population is wasteful of resources, not cost-effective, and will not benefit those who truly stand to benefit. Focusing in on the most vulnerable individuals is where the power of digital health interventions lies. Machine learning outputs, along with the sets of distributed weighted algorithms (which are designed from the outset and based on behaviour change frameworks) are used in identifying target populations, the intervention design, and communications delivery.



MEMOTEXT and partners then generate intervention proposals for consideration based on the learnings and identification of nodes for behaviour or population trajectory change. For example, designing an intervention to prevent patients from switching to a high-cost therapy



| EVIDENCE | EMPATHY | EXPERIENCE |

may mean targeting them earlier on in their disease state and offering higher touch supports to those labelled most at risk by the model. Other examples include:

- Connecting patients predicted to not adhere or comply with their medication with a pharmacist
- Initiating lifestyle improvement and behaviour change programs for those flagged as at risk of developing diabetes
- Enrollment in a management program for patients with schizophrenia or psychosis predicted to relapse or end up in the ER
- Support worker notification for clients predicted to attempt suicide

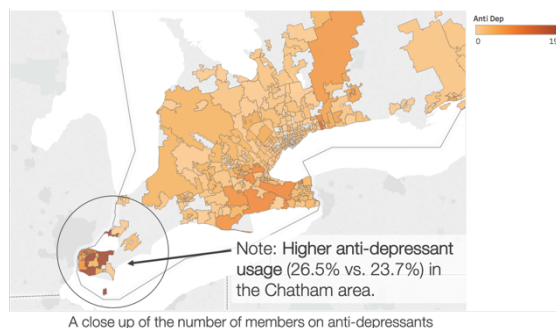


Figure 4. Mapping patients by features such as language, socioeconomic status, average income, number of available providers or anti-depressant use.

It is important at this stage to design the intervention in such a way that actionable data is collected (Figure 4). Post-pilot study is too late to modify data that has already been collected. MEMOTEXT specializes in designing digital interventions that seamlessly integrate with data analysis at every stage to evaluate results and validate the success of the program.

5. Deployment of model and Intervention

Once the intervention is designed and data collection methods along with KPIs are agreed upon, it is deployed alongside the model using implementation and behaviour change best practices. Machine learning outputs can be used

to improve and personalize the intervention as it's being executed. Data from the intervention is constantly fed back into the model for real-time effects on subsequent communications, program content distribution, and supports offered to the participants. Deploying the model alongside the intervention is one tool MEMOTEXT uses to personalize the intervention down to the individual patient, not just the patient population.

6. Operate, Optimize, and Iterate

Once the model is up and running it's important to keep it that way! Operation includes the maintenance tasks and checkpoints after project launch that facilitate the successful employment of the solution. As new features are imagined and developed, they can iteratively be included to further optimize the model. Additionally, as more and more data points are collected, the model will likely need to be re-trained and a new algorithm selected to ensure it is performing at the highest capacity on the most relevant data while being mindful to avoid overfitting or overtraining it.

Conclusion

The process of designing, implementing, and operating data-driven digital health interventions is complex and requires a breadth of technical and domain-specific knowledge along with years of experience. Understanding the technical aspects of machine learning is simply not enough for healthcare-based projects to succeed as it is a unique industry with its own nuances and complexities. MEMOTEXT continues to evolve its methodology, iterate on its platform and leverages real-world experiences to ensure project success, value added for stakeholders, and improved health outcomes for patients.

To learn more about our interventions, ask for our most recent MEMOTEXT Lit Review.