Data Transformation: Through the eyes of our Working Example

Our Story

At AAVAiL you are taking part in a 9:00 AM weekly playback meeting with teams from different parts of the company. A woman who works in the marketing department asks a question.

"Is there any way we can quickly analyze the feedback we get from our customers about the service?"

You know how to do natural language processing with Python so you raise your hand and speak up. It turns out that marketing wants to use those customer comments to automatically generate recommendations of content, as well as to predict customer churn.

You know the basics of data transformation, but now you get to use that knowledge in a more applied setting. An important aspect of applying data transformations is making sure you apply them correctly in your AI workflow. This in turn generally involves iteration.

Feature engineering is about the selection of data "features" that enhance the performance of your model. Another useful aspect of feature engineering is the augmentation of your original matrix with *new* features that often contain domain specific insight. All of the transforms for feature engineering we will touch on are meant to be compared with each other. From a software engineering perspective this will mean working with pipelines, because they are useful to compare variants of the workflow.

THE DESIGN THINKING PROCESS

Transforming data and feature engineering are processes that can take place at any stage of the design thinking process. In this case, AAVAiL has completed most of the data gathering work associated with the *Empathize* phase and is moving on to the *Define* phase where information and data are being sorted and organized in order to define the challenges that AAVAiL is facing.

This is one point at which we can start to incorporate domain experts to help us to translate their insights into usable features in our data. Playbacks, interviews, and data understanding exercises are all important tools we would use to take advantage of the knowledge of domain experts.