Brainstorming

A diagram of a process flow

Description automatically generated

1. Gather – already done

2. Cleaning and inspecting data to better understand it

3. Feature engineering to allow an algorithm to leverage the data in a suitable form (converting data to numerical vectors

Look at the data - structured, labelled. Can use supervised learning.

Supervised learning - train a model to predict the values of those labels based on various features of the data points.

* iterative optimization algorithm such as gradient descent.
* The training algorithm starts with a basic model and gradually improves it by adjusting various internal parameters (coefficients) during each training iteration.
* The result of this process is a trained model that you can use to make predictions on new data.

Types of supervised learning –

Classification - the act of training an algorithm to predict a dependent variable that is categorical (belonging to a discrete, finite set of values).

* Our dependent variable is the style preference. We want our model to be able to predict which 1 of 4 styles a person would choose.
  + At the beginning, we don’t know which variables this style preference depends on, which is what we are going to use ML to help us determine.
  + Dependent variable is a set of discrete values – Yes – that is what we have.
* Once satisfied with the algorithm’s performance, we use that model to make predictions
* Common case binary classification - resulting model will make a prediction that a given item belongs to one of two groups.
* Multiclass classification – classifying items into more than 2 categories

Regression – the act of training an algorithm to predict a continuous variable (a real number)

* Rather than predicting a category, predict a number along a line
* Predicting total sales, predicting child height based on parent height, # viewers on a show
* This doesn’t seem like what we need to use for what we are trying to train, nor with the collected data

Recommendation –

* Most intuitive applications of advanced analytics
* Through ratings or observed behavior, make recommendations based on what users may like based on similarities between users or items
  + Absolutely what we are trying to do with the dataset
  + We don’t know if it’s the person’s age, height, body type, color preference, etc. that matters. We do know things like if they returned an item, and how they rated the item
  + What we are trying to do is figure out how to recommend the right style of outfit, predictively.

Graph Analysis –

* Study of structures –
  + Vertices – objects
  + Edges – relationships between those objects
* Anomaly detection –
  + How networks of individuals connect, and outliers/anomalies can be flagged for manual analysis
* Classification –
  + given facts about vertices in a matrix, can classify other vertices to their connection to their original node –
* Recommendation –
  + Example: Google’s original web recommendation algorithm, PageRank, is a graph algorithm that analyzes website relationships
    - in order to rank the importance of web pages. For example, a web page that has a lot of links to it is ranked as more important than one with no links to it.