**Tech Insights – Clustering**

I’ve chosen the “The Predictive Power Of Social Media” article as it sounded interesting to learn about and relevant to a lot of current trends.

It’s interesting that the idea is to find out the question for the “unknown unknowns” or those that you don’t know how they are connected and are using clustering to try to find the question. It does mention other methods of classification but focuses on how clustering can help us to expand our thoughts to what other questions we maybe SHOULD ask regarding the data that wouldn’t be obvious but may be useful.

It’s wording/defining of clustering as not just the relationship between two points of data but rather how likely the two points are to be close for prediction to see that relationship. I feel like we often talk about the distances but this perspective is to think moreso in the “will these continue to be close” as if the answer is yes then what does it mean? What does that correlation mean to what we want to learn? And if the answer is yes how can we use that knowledge to predict things based on the data.

However it also clearly mentions that we don’t inherently by clustering know WHAT that correlation means when we see it, that’s where we need to dig deeper to see why we might be getting that result for example.

For example recognizing a new trend to capitalize on it. So maybe we notice that recently people who bought pretzels had a correlation to also getting herb and garlic cream cheese. Is it a coincidence? Is it a new snacking trend?(I think so! Maybe we test this out!) and so maybe we want as a grocery store to have one of those baskets for goods with pretzels located next to the cream cheese to see if we can grow this trend or capitalize on it.

Or for Social media more specifically to the article maybe when clustering we are seeing more instances of these words together, so perhaps people mentioning to their friends that it’s been their go to snack lately, or maybe there was a tik tok oddity showing an experiment with those items skewing it etc. These are the questions we might want to ask to see how we could use this information.

Their example about music trends in social media notes “there is an immensely strong correlation (r2 of .94) between record sales and social media activity. There appears to be a two-week lag between activity and sales. “ can we use that to get new artists names out there and be prepared?

There’s a lot to think about overall about how clustering can help us or make us open our minds.