1.2 Learning Response: Predictive Analytics Use Cases

Jason “Scott” Person

20250326

Response Questions

1. **From Siegel:** Select 5 key ideas from his introduction. For each, provide a well-phrased heading, summarize the idea, and then reflect on its implications in approximately 150-200 words.

My strategy for this question is to pick the ideas that I found most interesting. They are key ideas, but they may not be THE key ideas. I’m trying to get deeper into the material than just listing the obvious points.

1. Predictive Analytics is Completely Different Than Forecasting

On page 16 of the introduction, Siegel makes the point that forecasting is done on a macro level whereas predictive analytics (PA) makes predictions at the individual event level.

This follows very closely the ideas presented by McCormick in his videos. In the past it was not possible for an organization to record each individual event and then “mine” that data set for insights into the expected outcome of future individual events. The capabilities did not exist. Instead, aggregate trends would be compiled, and future expected aggregates would be derived from them. For example, a company having experienced 10% growth over each of the past three years might with reasonable confidence forecast that growth this year will be 10%. Enter modern PA techniques and the computing power that enable them. The same company can now analyze each individual transaction of potentially billions of transactions and predict the behavior of individual customers. Thus, in addition to knowing the general level of the resources for the year (from the forecast), the company also knows who will be buying their products, where they will be buying them, and how many they will buy at a time.

Another implication is that there’s plenty of opportunity here. The benchmark is no longer perfection, but merely improvement and comparative advantage, and the scope of appropriate problems is much broader. The power of scale is multiplicative – small enhancements in predictive capability and thus organizational efficiency when applied across large systems can yield substantial aggregate benefits. Because the potential gains are subtle and themselves difficult to recognize a culture of experimentation is essential. Testing, iteration, careful measurement, and frequent failure are drivers for comparative advantage.

1. Accuracy is not Required

Page 13: “Predictions need not be accurate to score big value.” (Seigel)

Even small gains in predictive success can yield benefit.

In a competitive market, an organization can derive extreme value from very small benefits. The key factor here is that an organization does not need to be perfect or even close, it just needs to be better at predicting than its competitors. Being incrementally better (more accurate) at PA can translate into meaningful strategic superiority. Conversely, a failure to at least maintain parity with competitors could present existential challenges.

1. Predictive Analytics is for Organizations

Page 18: Siegel states that “With only a few striking exceptions, we find that organizations, rather than individuals, benefit from employing PA.

This is fascinating and there are a couple of drivers. First, organizations have the data scale necessary to feed the PA process. Second, organizations have the problem scale that make the insights provided by PA worth the cost.

To be successful with PA, we need a significant amount of data. Further, each individual prediction need not be particularly accurate (see above). This leads to requirements of scale on both the inputs as well as the application of the output model(s). Organizations would generally have more of these in comparison to individuals.

Siegel also points out that the inherent inefficiency of organizations, and implicitly the relative efficiency of individuals, gives them more opportunities to apply PA. All of this driven by profit motive for the private sector and in government/non-profits by the desire to do more in the context of limited resources.

The above is interesting when compared to recent developments in generative AI. Here, organizations are struggling to apply it whereas individuals are finding novel implementations daily.

1. Learning is Hard

Page 20: “It turns out that learning – generalizing from a list of examples, be it a long list or a short one – is more than just challenging. It’s a philosophically deep dilemma.” (Siegel)

Note: This is where the professor on reading this knows that the student didn’t just pick the easy key ideas on which to pontificate in a contrived academic voice.

Machine learning fundamentally is taking a large number of specific events (discussed above), creating a general model, and then using that model to predict future specific events. Siegel points out that this isn’t just data “science”, it’s also data “art”. He goes on to say “The machine actually learns more about your next likely action by studying *others* than by studying *you*.” This is the last thought in the introduction and is obviously intended to entice the reader to eagerly consume the rest of the text. It’s working. I’m going to buy the book. I believe what Seigel is alluding to here is that for other than trivial edge cases there are more events and associated data for the action of “others” than there are for any single individual. Also a specific action by an individual is circumstantial. The prospective actions of that same individual would tend to mirror the population’s actions as predicted by the population’s historical data.

1. Geeks ~~Rule~~ Power the World

Page 18: “The source, the energy that makes it [here speaking of PA], is Geek Power! I speak of the enthusiasm of technical practitioners.”

Note: here I’m going to switch to first person primarily because I consider myself an expert at being a geek, but also because I think I’ve been taking this assignment too seriously. Note to the prof: correct me if I’m wrong.

This was my favorite concept from the introduction. It encapsulates the extreme (!!!) fortune that I’ve experienced in my career which is being able to do something so enjoyable while adding outsized value enabling outsized compensation. In other words, I’ve been able to make a lot of money doing fun stuff. Siegel points to the enjoyment as being the defining indicator of geekdom. That’s a good point, but we can’t lose sight of the fact that the opportunity is provided by the value added. Above I discussed the outsized impacts of small improvements resulting from imperfect models. This is the driver of the value add. Software and more specifically models scale well. The marginal cost of applying a model is zero. Provided the model has even a small positive net value on each application, the resulting upside is only limited by the number of predictions that can be made. Even small things done a bunch of times aggregate to big changes – in this case big profits.

The challenge for the geek is that the last 20% of doing geek stuff is often hard and not as fun. We have to focus on getting things done in order for the organization to benefit. Thankfully (haha) we have managers and coworkers willing to provide motivation and assistance with the mundane.

1. **From both Siegel and SeattleDataGuy:** Compose a list of 12 to 15 predictive analytics use-cases — real-world problems for which predictive analytics can be a useful tool. These can be phrased in in this pattern: "To predict which customers will be most likely to respond to an offer."
2. Predict the best time to buy a plane ticket based on expected future fare changes.
3. Predict stock market behavior based on historical events. Note, being intentionally vague here because it could be anything. I recently read a book on this topic: The Man Who Solved the Market (Zuckerman), which is about Jim Simons and his firm Renaissance Capital. Recommended.
4. Predict dating compatibility of singles
5. Predict thoughts of an individual based on real-time MRI scans
6. Predict driver attention level
7. Identification of breast cancers in radiology scans. Anecdote: my wife had a mammogram this week. The results were in before she got to the car. I told her that they were undoubtedly using models to compare previous scans with the new ones.
8. Detecting patient schizophrenia based on transcripts (!)
9. Identify lying based on visual indicators
10. Customer sentiment detection based on social media
11. Predict the best retail location
12. Predict patient readmission rates (this one is in both sources)
13. General and various fraud detection across many industries
14. **From your additional research:** In APA or MLA format, provide the information for your source, including a link if it's an online source.   
    Then, in approximately 250 words, summarize the predictive analytics use case, answering the questions:
    * What is the purpose or problem?
    * What benefits does it offer? To whom?
    * If available, look for evidence to answer: Approximately how big of a difference does it make? What kind and/or percentage of improved results, etc.?
    * Add your own reflection to tie it up.

Source:

Xu, L., Gholami, S., McCarthy, S., Dilkina, B., Plumptre, A., Tambe, M., Singh, R., Nsubuga, M., Mabonga, J., Driciru, M., Wanyama, F., Rwetsiba, A., Okello, T., & Enyel, E. (2019). *Stay ahead of poachers: Illegal wildlife poaching prediction and patrol planning under uncertainty with field test evaluations*. arXiv. <https://arxiv.org/abs/1903.06669>

The authors developed enhancements to the Protection Assistant for Wildlife Security (PAWS), which is an ML pipeline that predicts areas at high risk for poaching activity. Their enhancements mitigated data quality issues and improved detection of snares by 30%. PAWS generally is used to optimize patrol routes. Beneficiaries include wildlife management officials as well as obviously the animals themselves.

When looking for a response to this question I intentionally sought out something obscure. These models were created using data from Uganda and Cambodia then applied in the same regions. It would be stereotyping to assume that the data from game managers in developing countries is bad, but that is indeed the case. The managers concerns are life or death for both them and the wildlife. One focus of the study is how to deal with the “unreliable and imbalanced data”. An example, 99.6% of the labels were negative indicating no illegal activity. Was there truly no activity, or was it just well concealed? The fact that the authors were able to have a positive impact and get quality results from their models is encouraging and inspiring.