Machine Learning Analyst Case Study

At its core, our team's purpose is to extract insights from data. This effort spans many disciplines (investing, user experience, content generation, etc.). The usefulness of those insights will be dictated by many factors – data availability and quality, modeling acumen, process controls, etc. – but perhaps the most important criterion is effective communication.

In this spirit, we would like to offer you an opportunity to present your approach to tackling a stylized business challenge. Please take some time over a week-long span (ideally about 8-to-10 hours, though we're happy to offer additional time upon request) to assemble a **presentation** on the topics below. Once your presentation is ready, we'll coordinate for you to deliver it to a panel of stakeholders across our firm. Feel free to use either PowerPoint or Google Slides, and plan for the final presentation to take about **45-60 minutes**.

(For this exercise, please focus on communicating the business value of your approach rather than deliberating on specific technical details. We really want to see how effective you are at making key connections and explaining difficult topics to individuals that may not have technical backgrounds!)

Background:

Every great modeling solution needs an equally great business problem. At The Motley Fool, our purpose is to help each of our subscribers – a.k.a. our members – on their paths toward becoming *smarter*, *happier* and *richer*. This definition of "success" will be different for each member. Here are some examples to illustrate:

- Members want different experiences from their subscriptions (e.g. a simple way to get into the game of investing, a full portfolio management solution, a pure "hobbyist" experience)
- Members have different investment objectives (e.g. high growth vs. a steady dividend stream).
- Members have different degrees of risk aversion (e.g. risk seeking vs. risk averse members)
- Members learn in different ways (e.g. reading articles vs. watching a live streaming segment)
- Members derive satisfaction in different ways (e.g. some members may view emailed articles as a nuisance while others appreciate emailed articles)

Objective:

For this case study, please take some time to design an approach for moving our subscribers along their respective paths toward "success" using a data-driven approach. You may apply any model framework you like, but here's a sample single-period structure to capture dynamics from time *0* to time *t*:

- Member success over some future time period τ, $\mathbf{Y}_{\tau}$, is represented by an N_0 x K matrix, where N_0 is the number of members and K is the number of success variables (which can be observed or latent).

- The member feature matrix at a starting time t_0, $\mathbf{X_0}$, is represented by an N_0 x Z_0 matrix, where Z_0 is the number of candidate member characteristics (which can be observed or latent).
- The content matrix, $\mathbf{C_0}$, is represented by an M_0 x L matrix, where M_0 represents the number of unique content pieces and L represents the range of content metadata characteristics.
- The member-content interaction matrix, $\mathbf{I_0}$, represented by an N_0 x M_0 matrix that describes the targeted mapping between members and content offerings.

Outputs:

Please prepare a presentation on the following:

1. How might we arrive at the optimal expressions for **Y**, **X**, **C**, and **I** from the above framework (or, if you applied a different framework, how would you arrive at the corresponding optima)?
    a. Consider the data – which data sources might be helpful in the modeling effort? Think big! Don't assume there are constraints to acquiring useful data, but rather strive to identify the most useful input data sets.
    b. Define the target(s) – how might we measure "success?" Feel free to consider a comprehensive approach, or to design incremental outcome variables instead. Your outcome variable(s) can be binary, real-number-valued, probabilistic, etc. You get to decide.
    c. Discuss the features – which member characteristics might be helpful in tracking and/or influencing success?
    d. Discuss the decisioning variables and interventions. Which content metadata tags could prove to be most important? How might we determine this?
2. Consider the context and potential confounders.
    a. Which confounders might be especially important to consider for this business problem?
    b. Which might be especially unique to our challenge?
    c. How would you go about modeling around these confounders to arrive at true causal relationships?
3. Propose the modeling effort. Which techniques would you try out first? How would you proceed through the testing and model selection phase? (No need to get super detailed here!)
4. What would you suggest for next steps? Imagine that we would like you to start tomorrow!