Article from

CompAct

October 2018

Issue 58

Alternative Target Operating Model for Your Predictive Analytics Function

By David Schraub

The use of predictive analytics to drive the sales and marketing strategies of insurance companies is a trend on the rise. Predictive analytics models used to determine measures such as propensity to buy, etc., is a common way in which predictive analytics is used to inform the sales and marketing strategies of insurers.

The starting point for any predictive analytics initiative is to clearly identify the business need or problem. This will, amongst other things, drive the choice of the model used.
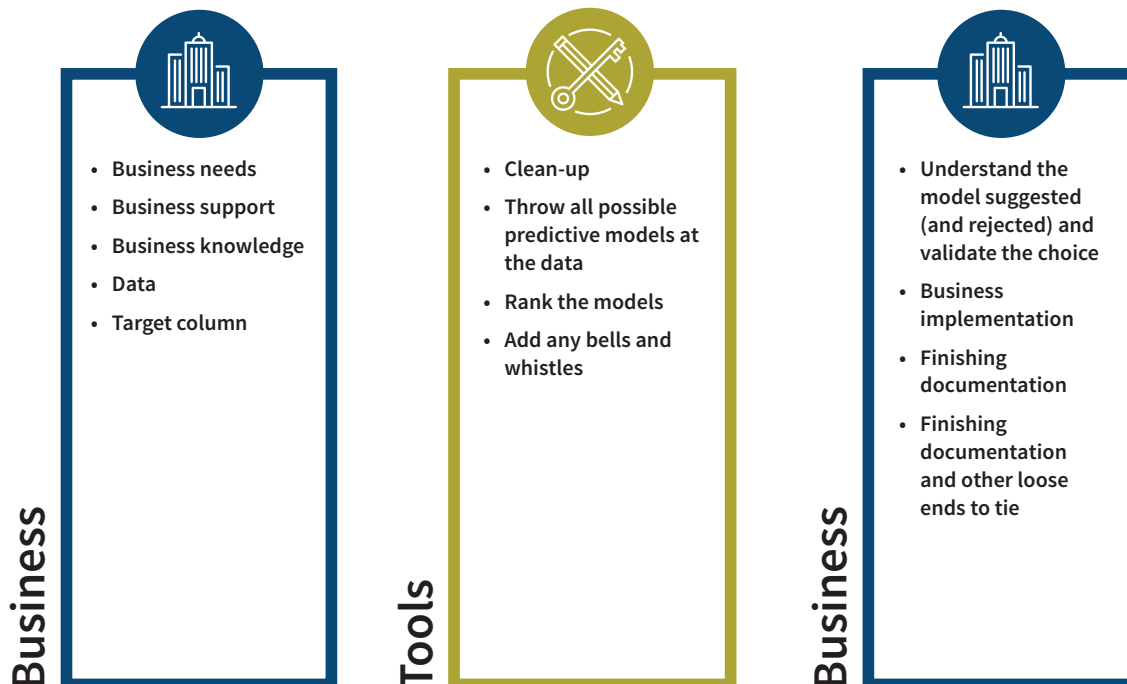
Below is a list of the common steps to be followed for any predictive analytics task:

1. Integrate this predictive analytic endeavor in a business context[1]
2. Define the data needed
3. Clean the data
4. Choose model type and build the model
5. Test Model
6. Implement

The usual target operating model for the predictive analytics function is to build a full predictive analytics team. One alternative would be to leverage some tools that can automate some of the time consuming processes. Examples of such software platforms are DataRobot, Alteryx and RapidMIner (and WTW Emblem for property and casualty (P&C) pricing). The goal of such a tool is to lighten[2] the need for quants on the predictive analytics team by facilitating the model building process.

This tool will not help you with business problems[3] and get data for you. This will not replace business knowledge or data dictionary and will not solve the garbage-in-garbage-out issue. You are on your own here.

What the tool does is to help automate simple cleaning (recoding null values, creating manageable numbers of buckets if there are too many possible discrete values, using the word cloud function to turn text into variables, detecting outliers, etc.),

Business

- Business needs
- Business support
- Business knowledge
- Data
- Target column

Tools

- Clean-up
- Throw all possible predictive models at the data
- Rank the models
- Add any bells and whistles

Business

- Understand the model suggested (and rejected) and validate the choice
- Business implementation
- Finishing documentation
- Finishing documentation and other loose ends to tie

build all possible models on a subset of the data (GLM, random forest, gradient boosted trees classifier with early stopping, etc.), test and validate on another subset of the data, and rank all the models by their predictive powers. It adds all the possible bells and whistles for you: Open code visible, technical documentation in Microsoft Word, graphical representation of the predictive force of each variable (with dependence analysis) and many more that I didn't list here.[4]

From there, the business needs to pick it back up. Using the documentation, individual scenarios and other graphs, you need to open the black box and understand the model chosen to fully own it. You can leverage this to more easily explain your model to other stakeholders and get buy-in. You still need to clean up a few loose ends on the documentation and on other areas in order to get to the implementation.

CONCLUSION

No, this tool doesn't bring you coffee. But it could help reduce costs and banish the fear of the investment involved in building a full predictive analytics function with full-time employees. ∎

David Schraub, FSA, CERA, AQ, MAAA, is a staff actuary for the SOA. He can be contacted at dschraub@soa.org.

ENDNOTES

1 The predictive analytic component exists to solve a business question (e.g. optimize possible prospect, detect potential fraud, or retain customers) in a business context (project funding, timeline…)

2 You still need predictive analytics quants on your team, in order to understand what the black box just did. Pressing a magic button without deep knowledge could be potentially problematic.

3 I attended a presentation from a provider of such tool. One of the examples they use on their site and in a live demo is about loan performance. Their choice of target column is a Yes/No column representing the loan default, instead of capturing a profitability metric. The speaker acknowledged off-line that this misses all the high-yield profitable loans that default and wrongly includes the low yield performing loans … but makes the inner working of the tool easier to explain.

4 I also admit didn't grasp the usage of all the bells and whistles that were presented during the tool provider session I attended.