INSIGHTS · TOOLS AND THE FUTURE

Predictive Insight and Forecasting Customer Behaviour

How organisations forecast what customers will do, from sales forecasts and new product adoption to churn and customer value, what forecasting competitions and research show about which methods work, how to judge a forecast, and why predicting is not the same as knowing what to do.

ModQu research article · 2,008 words · 24 references (APA 7th edition)

Much of the value of insight lies in looking ahead. Organisations want to know how much they will sell, which customers will leave, which will buy again, how a new product will spread and what customers will be worth over time. Predictive insight is knowledge about customers that supports such forecasts. The field has advanced rapidly with the growth of data and machine learning, but some of its most useful lessons come from decades of research on forecasting accuracy. This article summarises what is known about forecasting customer behaviour, how to judge a forecast, and how to turn predictions into good decisions.

Prediction and explanation are different goals

Shmueli (2010) distinguished explanatory modelling, which aims to test causal theories, from predictive modelling, which aims to forecast new observations accurately, and showed that the best model for one purpose is often not the best for the other. Breiman (2001) described two cultures in statistics: one that assumes a model of how the data were generated, and one that treats the process as unknown and judges models by their predictive accuracy. Machine learning belongs mainly to the second.

For insight teams, the distinction matters because a model that predicts well does not necessarily explain why. A churn model may predict accurately using variables that are symptoms of dissatisfaction rather than causes. Kleinberg et al. (2015) argued that many important decisions depend on prediction rather than causal understanding, for example deciding whom to prioritise, but that the two kinds of problem should not be confused.

What forecasting research shows

Simple methods are hard to beat

Forecasting competitions have tested many methods on large numbers of real time series. Makridakis and Hibon (2000) reported the results of the M3 competition, in which more than 20 methods were compared on 3,003 series, and found that statistically sophisticated methods did not necessarily produce more accurate forecasts than simpler ones, and that combining forecasts from several methods improved accuracy. Makridakis et al. (2020) reported the M4 competition,

with 100,000 series, and found that combinations of methods and hybrid approaches that blended statistical and machine learning methods performed best, while pure machine learning methods generally did not outperform simple statistical benchmarks on those data.

Hyndman and Athanasopoulos (2021) recommended that any forecasting method be compared with simple benchmarks, such as the last observed value or the value at the same time last year. A method that cannot beat these benchmarks adds complexity without value. Armstrong (2001) compiled evidence-based principles for forecasting, including using simple methods where data are limited, combining forecasts, and being conservative when uncertainty is high.

Judgement helps and hurts

Most organisations adjust statistical forecasts by judgement. Fildes et al. (2009) studied tens of thousands of forecasts in supply-chain companies and found that judgemental adjustments were very common. Large adjustments, often based on specific information about promotions or events, tended to improve accuracy, while small adjustments added little, and positive adjustments were often too optimistic. Their recommendation was to adjust only when there is specific, relevant information and to record the reasons.

Research comparing expert judgement with statistical prediction more generally has found that formulas often match or beat experts. Grove et al. (2000) found in a meta-analysis that mechanical prediction was on average about 10% more accurate than clinical judgement. Tetlock (2005) found that expert forecasts of political and economic events were often little better than simple rules.

Forecasting customer behaviour

Who is still a customer, and what will they buy?

In many businesses, customers can stop buying without saying so. Schmittlein et al. (1987) proposed a model that uses each customer's history of purchases, how many and how recently, to estimate the probability that the customer is still active and how many purchases they will make in future. Fader et al. (2005a) proposed a simpler model with similar accuracy that is easier to apply, and Fader et al. (2005b) showed how the recency, frequency and monetary value of past purchases can be used to estimate future customer value. These models rest on regularities in repeat buying that Ehrenberg (1988) documented across many product categories.

Churn

In subscription businesses, churn models predict which customers are likely to cancel. Neslin et al. (2006) organised a tournament in which teams built churn models for the same telecommunications data, and found that the accuracy of models differed considerably, that the choice of method and the variables used both mattered, and that better models could generate substantially more profit from retention campaigns.

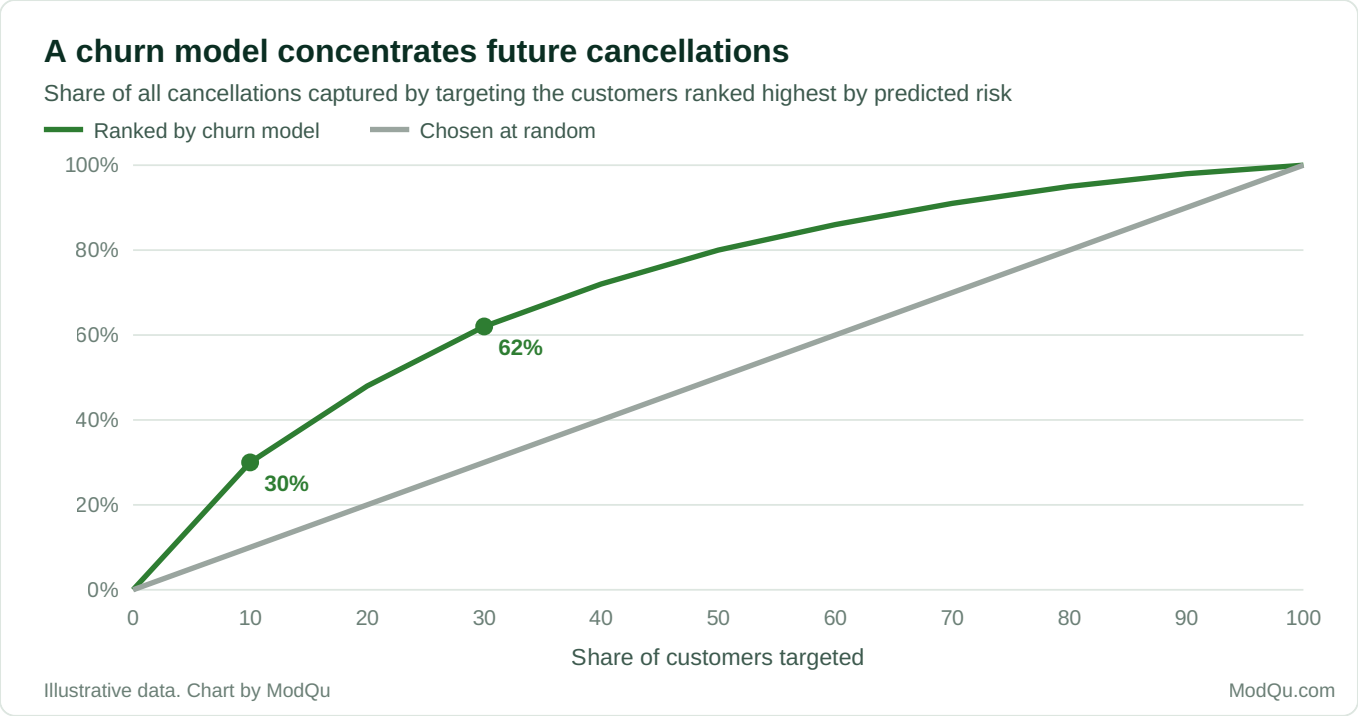


Figure 1. A churn model concentrates future cancellations among the customers it ranks highest. Illustrative data, chart by ModQu.

But predicting who will leave is not the same as knowing whom to target. Ascarza (2018) analysed retention experiments and found that the customers at highest risk of leaving were not necessarily the ones most likely to be persuaded to stay by a retention offer. Targeting customers by their predicted response to the intervention, rather than by their predicted risk, was more effective. Lemmens and Gupta (2020) proposed targeting retention campaigns by their expected effect on profit, taking account of both the effect of the offer and its cost.

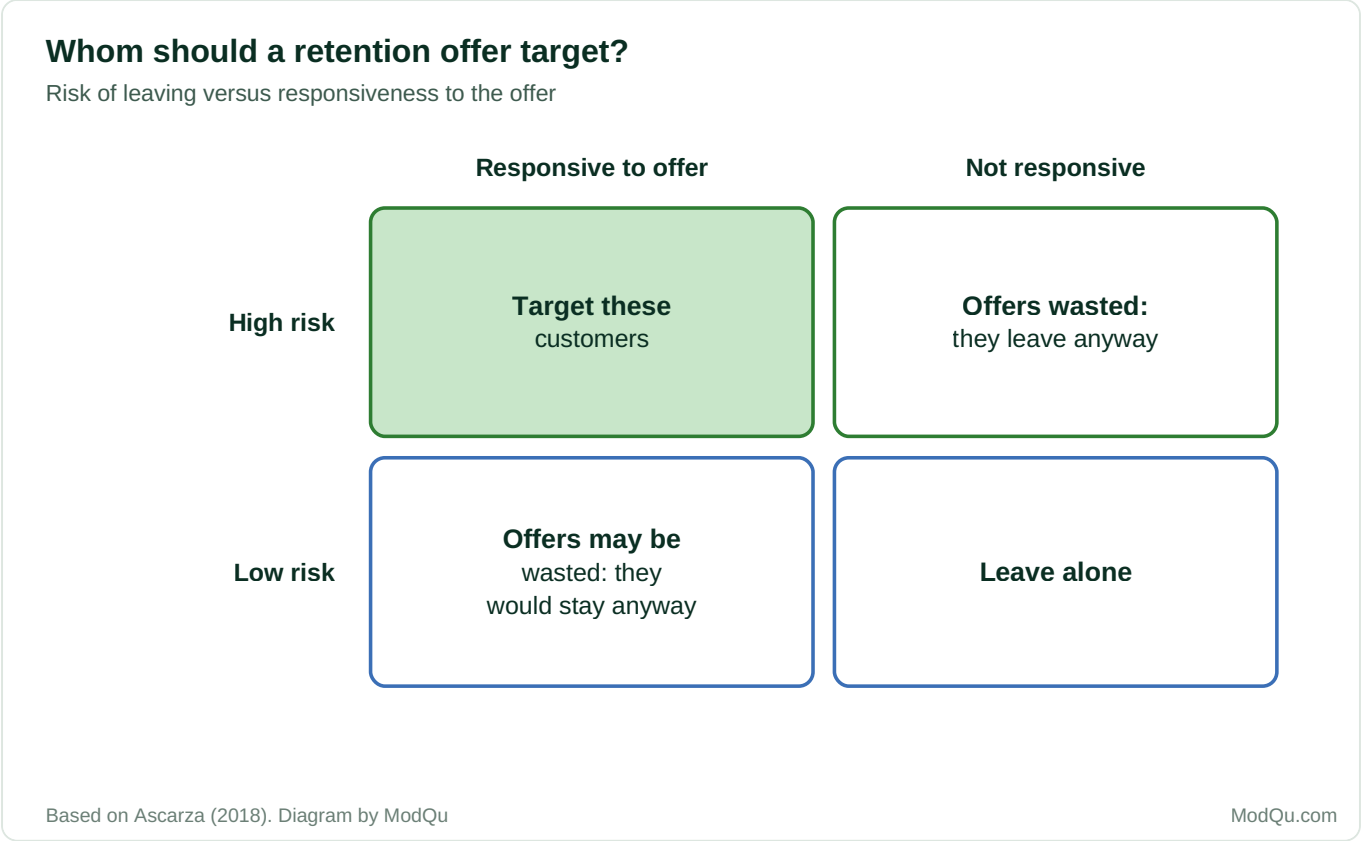


Figure 2. Whom to target depends on whether an offer will change behaviour, not only on risk. Diagram by ModQu, based on Ascarza (2018).

New products

Forecasting demand for new products is harder, because there is no sales history. Bass (1969) proposed a model of how new durable products spread through a population, driven by innovators who adopt independently and imitators who adopt because others have. The model remains widely used, often with parameters estimated from similar past products.

Purchase intentions are a common input, but they need adjustment. Juster (1966) found that asking people for the probability that they would buy predicted purchases better than asking whether they intended to buy. Morwitz et al. (2007) reviewed studies linking intentions to sales and found that intentions predicted sales more accurately for existing products than for new ones, for durable goods than for non-durables, and over short time horizons than long ones.

New data sources

Online behaviour can add predictive information. Goel et al. (2010) found that the volume of web searches for films, video games and songs before release predicted their later success, although in many cases search data added relatively little to simpler predictions based on other publicly available information. Lazer et al. (2014) showed how a model that estimated influenza levels from search data became inaccurate over time as search behaviour changed, a warning that new data sources can be unstable.

Common pitfalls

Several mistakes recur in predictive work on customers.

Leakage. A model is accidentally given information that would not be available at the time of the prediction. A churn model that includes "number of calls to the cancellation line" will appear highly accurate, because it uses the act of cancelling to predict cancellation. Such models perform well in testing and poorly in use. The remedy is to build each prediction only from data available before the date on which it would have been made.

Overfitting. Flexible models can fit the quirks of past data rather than the underlying pattern. A model with hundreds of variables, tuned to one year of data, may predict the next year poorly. Testing on later periods, keeping models as simple as the problem allows, and comparing with simple benchmarks guard against this.

Ignoring base rates. When the event being predicted is rare, even an accurate model will produce many false alarms. If 2% of customers cancel each month, a model that correctly flags 80% of cancellers and wrongly flags 10% of stayers will produce roughly six false alarms for every true one. Decision makers who see a list of "at-risk" customers should understand that most of them will not leave.

Optimising the wrong target. A model is built to predict what is easy to measure, such as clicks or short-term purchases, when the decision depends on something else, such as long-term value. A model that perfectly predicts the wrong outcome can lead to worse decisions than no model at all.

Static models in a changing world. Customer behaviour shifts with prices, competitors, technology and the organisation's own actions. Models need monitoring and periodic retraining, and their assumptions need review when the market changes.

Judging a forecast

A forecast should be judged against what happened and against simple alternatives.

Accuracy against benchmarks. Compare errors with those of naive forecasts, such as last year's value.

Calibration. For probabilistic forecasts, such as the probability that a customer will cancel, check that events predicted with, say, 30% probability happen about 30% of the time. Gneiting and Raftery (2007) described scoring rules that reward forecasts for being both calibrated and sharp.

Out-of-sample testing. Judge a model on data not used to build it, preferably from a later period.

Stability. Monitor accuracy over time, since relationships change. A model that was accurate when built may drift as customers, products and markets change, and a gradual decline in accuracy is often the first sign that an underlying insight has stopped being true.

Value. Judge a model by the improvement in decisions it produces, such as the extra profit from a better-targeted campaign, not by statistical fit alone.

From prediction to decision

A prediction becomes valuable only when it changes a decision. Several practices help:

- **Express forecasts as ranges**, not single numbers, so that decision makers can plan for uncertainty.
- **Combine forecasts** from different methods and sources.
- **Separate prediction from intervention.** Use experiments to learn which actions change outcomes, and target actions where they have most effect.
- **Record judgemental adjustments** and their reasons, and review whether they improved accuracy.
- **Watch for feedback effects.** Acting on a forecast can change the outcome. A retention campaign aimed at predicted churners changes who churns, which affects how the model should be evaluated and retrained.

Building a forecasting practice

Organisations improve their forecasts most when forecasting is treated as a skill to be measured and developed. Useful practices include keeping a log of forecasts with the date, the method, any judgemental adjustment and the reason for it; comparing each forecast with the outcome when it is known; reviewing accuracy regularly, by method and by forecaster; and sharing what is learned. Tetlock and Gardner (2015) described how the most accurate forecasters in a large forecasting tournament kept careful score of their predictions, reviewed their errors and updated their methods, and argued that this habit of measured feedback is one of the main reasons some forecasters improve while others do not. Without such a record, organisations cannot tell which methods or people forecast well, and the same mistakes recur.

An example

Consider a hypothetical streaming service that builds a churn model from viewing and payment data. The model ranks customers well: the 20% it judges most at risk include about half of those who later cancel. The company initially sends a discount offer to this high-risk group.

A controlled experiment then compares the offer with no offer across all risk levels. It finds that many of the highest-risk customers cancel anyway, while a group of medium-risk customers, who had recently reduced their viewing but still used the service weekly, respond strongly to the offer. The company retrains its targeting to predict response to the offer rather than risk alone, and the same budget retains noticeably more customers. The original model was accurate; it was simply answering a different question from the one that mattered for the decision.

Key takeaways

Predicting customer behaviour and explaining it are different goals, and models that predict well may not explain why. Forecasting research shows that simple methods are hard to beat, that combining forecasts improves accuracy, and that judgemental adjustments help mainly when based on specific information. Established models forecast repeat purchasing and customer value, churn models can concentrate risk, and diffusion models and adjusted intentions help with new products. Forecasts should be judged against simple benchmarks, for calibration and out of sample. Predicting risk is not the same as knowing whom an action will help, so experiments are needed to turn predictions into good decisions.

References

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Kluwer Academic. https://doi.org/10.1007/978-0-306-47630-3

Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research, 55*(1), 80–98. https://doi.org/10.1509/jmr.16.0163

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science, 15*(5), 215–227. https://doi.org/10.1287/mnsc.15.5.215

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science, 16*(3), 199–231. https://doi.org/10.1214/ss/1009213726

Ehrenberg, A. S. C. (1988). *Repeat-buying: Facts, theory and applications* (2nd ed.). Charles Griffin.

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005a). "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Science, 24*(2), 275–284. https://doi.org/10.1287/mksc.1040.0098

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005b). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research, 42*(4), 415–430. https://doi.org/10.1509/jmkr.2005.42.4.415

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting, 25*(1), 3–23. https://doi.org/10.1016/j.ijforecast.2008.11.010

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association, 102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences, 107*(41), 17486–17490. https://doi.org/10.1073/pnas.1005962107

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.

Juster, F. T. (1966). Consumer buying intentions and purchase probability: An experiment in survey design. *Journal of the American Statistical Association, 61*(315), 658–696. https://doi.org/10.1080/01621459.1966.10480897

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review, 105*(5), 491–495. https://doi.org/10.1257/aer.p20151023

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Lemmens, A., & Gupta, S. (2020). Managing churn to maximize profits. *Marketing Science*, *39*(5), 956–973. https://doi.org/10.1287/mksc.2020.1229

Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476. https://doi.org/10.1016/S0169-2070(00)00057-1

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

Morwitz, V. G., Steckel, J. H., & Gupta, A. (2007). When do purchase intentions predict sales? *International Journal of Forecasting*, *23*(3), 347–364. https://doi.org/10.1016/j.ijforecast.2007.05.015

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, *43*(2), 204–211. https://doi.org/10.1509/jmkr.43.2.204

Schmittlein, D. C., Morrison, D. G., & Colombo, R. (1987). Counting your customers: Who are they and what will they do next? *Management Science*, *33*(1), 1–24. https://doi.org/10.1287/mnsc.33.1.1

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.1214/10-STS330

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. Crown.