

How to use predictive models

Quick tips on how to utilize individual-level predicted probabilities from statistical models in the field

[What is a predictive model?](#)

[Why use predictive models?](#)

[What types of predictive models exist?](#)

[How does one assess accuracy?](#)

[Predictive model and general targeting best practices, by program](#)

[Fundraising](#)

[Volunteer recruitment & prospecting](#)

[Persuasion](#)

[GOTV](#)

What is a predictive model?

Predictive modeling is the process of training a statistical (or machine learning) model to predict probabilities of outcomes, or future events, largely informed by a lot of additional, related data. In our case, many of our predictive models predict *something* about a registered voter.

Since we are dealing with probabilities, or the likelihood of some event occurring, the model is used to calculate individual predicted probabilities, or “scores.” These scores are usually on a scale of 0-100, or 0-1, with zero indicating no likelihood of the event occurring, and 100 or 1 indicating absolutely certainty of an event occurring.

However, even though these probabilities, or “scores,” are calculated on the individual level, it’s best to understand these scores in the aggregate rather than on a case by case level. For instance, if we have 1,000 voters, all with a turnout score of .60, or 60%, we would expect to see roughly 600 voters to go to the polls.

Why use predictive models?

Using predictive models is much more efficient for targeting voters for voter contact than targeting via vote history alone.

Remember, predictive models typically rely on algorithms to analyze hundreds, if not thousands, of data points related to voters in order to predict outcomes. These data typically include voting history, previous field IDs, demographics, consumer information, etc. A model that is well-validated can not only save countless hours, but can also help you target voters you would not have otherwise targeted when limited to Votebuilder queries alone.

What types of predictive models exist?

It's important to keep in mind that, with enough data, any outcome of interest can be modeled in order to produce individual-level predicted probabilities, or "scores." However, a few core predictive models are listed below:

- **Democratic support** predictive models estimate the likelihoods that voters will support the Democratic party
- **Candidate-specific support** predictive models estimate the likelihoods that voters will support a specific candidate
- **Turnout** predictive models estimate the likelihoods that voters will vote in a specific election
- **Persuasion** predictive models typically estimate the lift, or net new votes, voters could be responsible for if successfully persuaded to support a candidate.
- **GOTV** predictive models typically estimate the lift, or net new votes, voters could be responsible for if mobilized to vote.
- **Volunteer propensity** predictive models estimate the likelihoods that voters will be interested in volunteering
- **Phone contactability** predictive models estimate the likelihoods that voters are contactable via the telephone
- **Canvassability** predictive models estimate the likelihoods that voters are contactable at the door
- **Mail readership** predictive models estimate the likelihoods that voters read, or even recall, direct mail addressed to them
- **TV viewership** predictive models estimate the likelihoods that voters watch, or even recall, communications delivered via TV (n.b. think about broadcast vs. cable vs. satellite)
- **Digital consumption** predictive models estimate the likelihoods that voters consume, or even recall, digital ads or communication
- **Donation propensity** predictive models estimate the likelihoods that voters will donate

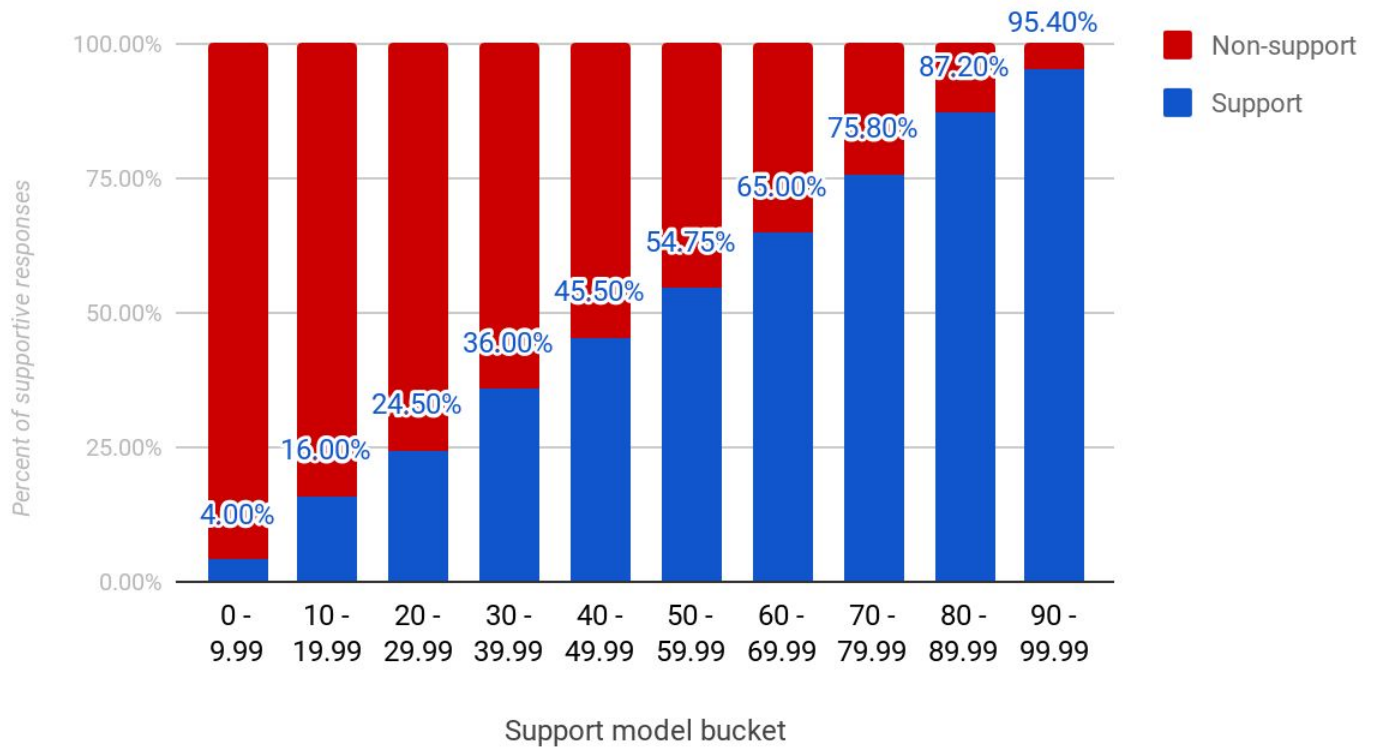
How does one assess accuracy?

There are two main ways to assess accuracy or prediction:

- **"Clinical" validation, or pre-release validation:** before a set of scores is released, the predictive model first needs to pass initial validation checks. Typically, this involves taking a subset of data and setting it to the side. This data includes dependent variables such as the outcomes, and other possibly helpful data, such as gender; age; marital status; etc., which are the independent variables. It is important to note that changes to independent variables can produce a different outcome, and therefore the outcomes are "dependent" on them. The rest of the data is used to fit, or train, the predictive model. This is where the algorithm helps do the analysis. An algorithm helps determine the relationship between the dependent variable (the outcome we're trying to predict), and the independent variables (all of the other data points describing voters). Once the model is trained, we bring back the data that was set to the side and see how well the model predicts the outcomes in that dataset.
- **"Field" validation, or post-release validation:** after a set of scores is released, you should monitor to make sure that the predictive model is still predictive of voters' responses or actions. Your Data Director in state can help set this reporting up for your state. Basically, you should code voters you've received survey responses for into their respective "buckets." For instance, if a voter's support model score is 87.5, they would fall into the "80.00 - 89.99" bucket. Then, calculate the percentage of voters

in that bucket who have provided a “supportive” response. Graph this data on a column chart, and look to see that the percent of supportive responses lines up with what we could expect. It’s probably a good idea to split these charts out by week or month, to guard against the predictive model starting off strong, but waning in predictive ability over time. See the example below:

[EXAMPLE] Support predictive model validation



Predictive model and general targeting best practices, by program

Fundraising

mode	best practices
doors (canvassing)	<ul style="list-style-type: none"> ask all supporters on the doors if they would like to make a small contribution to the campaign (keep contribution envelopes in your canvassing packets)
phones	<ul style="list-style-type: none"> ask all supporters on the phones if they would like to make a small contribution to the campaign and either: <ul style="list-style-type: none"> mark them as a “donor prospect” in VAN for quick follow up for your fundraising train phonebanking volunteers to process contributions while on the call
direct mail	<ul style="list-style-type: none"> if you have a donor prospecting model, you should either: <ul style="list-style-type: none"> work from a given budget for direct mail donor prospecting, and

	<ul style="list-style-type: none"> determine how many voters you can send mail to (by working down the score) and stay within that budget, or <ul style="list-style-type: none"> figure out a fundraising model score threshold to stay above and prepare to mail those voters you could also possibly include a generic mail readership model when building your universe to limit your direct mail universe to those who are likely to at least acknowledge receipt of direct mail in general (older voters moreso than younger voters typically)
digital ads	<ul style="list-style-type: none"> if you don't have a donor prospecting model, you can: <ul style="list-style-type: none"> load current or past donors into various social media platforms and find "lookalike audiences" to serve ads to calling them to donate if you have a donor prospecting model, you can: <ul style="list-style-type: none"> work from a given budget for digital donor prospecting, and determine how many voters you can serve up ads to (by working down the score) and and still stay within that budget, or figure out a fundraising model score threshold to stay above when figuring out what voters to match and target online with calls to action <ul style="list-style-type: none"> voters with a score ≥ 70 translates roughly to a set of voters who, together, 70% would likely agree to donate or actually donate you could also possibly include a generic digital consumption model when building your universe of potential donors to match to limit to: <ul style="list-style-type: none"> people who are likely to donate and likely to be active or consume media online

Volunteer recruitment & prospecting

mode	best practices
phones	<ul style="list-style-type: none"> for volunteer recruitment (recruiting people into shifts), it's usually best to start with people who have either: <ul style="list-style-type: none"> already completed a volunteer shift (reshifting) have indicated they would like to volunteer, but have not yet volunteered, and have not yet said they no longer want to volunteer once you've exhausted your list of people who have already indicated they're interested in volunteering, or those who have actually completed shifts, you can use a volunteer propensity predictive model to help find voters who would have a high likelihood of saying they're interested in volunteers you can possibly further cross this with a phone contactability predictive model to ensure you're more likely calling: <ul style="list-style-type: none"> people likely to say they want to volunteer, and people likely to be contactable via phone
digital ads	<ul style="list-style-type: none"> if you don't have a volunteer prospecting predictive model, you can: <ul style="list-style-type: none"> load current or past volunteers into various social media platforms and find "lookalike audiences" to serve ads to calling them to sign up to volunteer if you have a volunteer prospecting predictive model, you can: <ul style="list-style-type: none"> work from a given budget for digital volunteer prospecting, and determine how many voters you can serve up ads to (by working down the score) and still stay within that budget, or

	<ul style="list-style-type: none"> ○ figure out a volunteer prospecting model score threshold to stay above when figuring out what voters to match and target online with calls to action <ul style="list-style-type: none"> ■ voters with a score ≥ 70 translates roughly to a set of voters who, together, 70% would likely agree to volunteer or actually volunteer ● you could also possibly include a generic digital consumption model when building your universe of potential donors to match to limit to: <ul style="list-style-type: none"> ○ people who are likely to at least say yes to volunteering, and ○ likely to be active or consume media online
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Persuasion

mode	best practices
doors (canvassing)	<ul style="list-style-type: none"> ● with a persuasion model, and depending on how it's delivered (voters can be bucketed into rank ordered deciles or voters can be scored with an estimate of the number of net new votes they could produce if contacted), you can do one of two things: <ul style="list-style-type: none"> ○ figure out a good persuasion model score threshold to stay above, and generate a canvassing universe based on that ○ figure out how much door knocking capacity you will have over time. if you think your program can produce 10,000 knocks total, then you should figure out how far you can do go in the score to create a door knocking universe between 10,000 and 15,000 total doors. ● if you have access to a canvassability predictive model, you can further narrow, or target, your canvass universe to only include households that are more likely to result in a conversation at the door
phones	<ul style="list-style-type: none"> ● with a persuasion model, and depending on how it's delivered (voters can be bucketed into rank ordered deciles or voters can be scored with an estimate of the number of net new votes they could produce if contacted), you can do one of two things: <ul style="list-style-type: none"> ○ figure out a good persuasion model score threshold to stay above, and generate a phonebanking universe based on that ○ figure out how much phonebanking capacity you will have over time. if you think your program can produce 10,000 calls total, then you should figure out how far you can do go in the score to create a door knocking universe between 10,000 and 15,000 total unique phone numbers. ● if you have access to a phone contactability predictive model, you can further narrow, or target, your phone banking universe to only include households that are more likely to result in a conversation on the phone
direct mail	<ul style="list-style-type: none"> ● with a persuasion model, and depending on how it's delivered (voters can be bucketed into rank ordered deciles or voters can be scored with an estimate of the number of net new votes they could produce if contacted), you can do one of two things: <ul style="list-style-type: none"> ○ figure out a good persuasion model score threshold to stay above, and generate a direct mail universe based on that ○ figure out how much budget you have for a direct mail persuasion program and generate your direct mail universe based on those parameters (working down from highest household score down until

	<p>you hit the target number of mailboxes)</p> <ul style="list-style-type: none"> • if you have access to a mail readership predictive model, you can further narrow, or target, your direct mail universe to only include mailboxes that are more likely to actually read their mail
TV	<ul style="list-style-type: none"> • with a persuasion model, and depending on how it's delivered (voters can be bucketed into rank ordered deciles or voters can be scored with an estimate of the number of net new votes they could produce if contacted), you can do one of two things: <ul style="list-style-type: none"> ○ figure out a good persuasion model score threshold to stay above, and generate an "addressable" universe based on that ○ figure out how much budget you have for a TV persuasion buy and generate your "addressable" universe based on those parameters (working down from highest household score down until you hit the target number of mailboxes) • if you have access to a TV viewership predictive model, you can further narrow, or target, your universe to only include households chock full of target voters who actually consume TV
digital ads	<ul style="list-style-type: none"> • with a persuasion model, and depending on how it's delivered (voters can be bucketed into rank ordered deciles or voters can be scored with an estimate of the number of net new votes they could produce if contacted), you can do one of two things: <ul style="list-style-type: none"> ○ figure out a good persuasion model score threshold to stay above, and generate a digital voter contact universe based on that ○ figure out how much budget you have for a digital persuasion buy and generate your digital persuasion universe based on those parameters (working down from highest persuasion score down until you hit the target number of people) • if you have access to a digital consumption predictive model, you can further narrow, or target, your universe to only include people likely to consume digital media regularly

GOTV

mode	best practices
doors (canvassing)	<ul style="list-style-type: none"> • with a GOTV model, and depending on how it's delivered (typically, voters are scored with an estimate of the number of net new votes they could produce if contacted), you can do one of two things: <ul style="list-style-type: none"> ○ figure out a good GOTV model score threshold to stay above, and generate a canvassing universe based on that ○ figure out how much door knocking capacity you will have over time. if you think your program can produce 10,000 knocks total, then you should figure out how far you can go in the score to create a door knocking universe between 10,000 and 15,000 total doors. • if you have access to a canvassability predictive model, you can further narrow, or target, your canvass universe to only include households that are more likely to result in a conversation at the door
phones	<ul style="list-style-type: none"> • with a GOTV model, and depending on how it's delivered (typically, voters are scored with an estimate of the number of net new votes they could produce if

	<p>contacted), you can do one of two things:</p> <ul style="list-style-type: none"> ○ figure out a good GOTV model score threshold to stay above, and generate a phonebanking universe based on that ○ figure out how much phonebanking capacity you will have over time. if you think your program can produce 10,000 calls total, then you should figure out how far you can go in the score to create a door knocking universe between 10,000 and 15,000 total unique phone numbers. <ul style="list-style-type: none"> ● if you have access to a phone contactability predictive model, you can further narrow, or target, your phone banking universe to only include households that are more likely to result in a conversation on the phone
SMS	<ul style="list-style-type: none"> ● with a GOTV model, and depending on how it's delivered (typically, voters are scored with an estimate of the number of net new votes they could produce if contacted), you can do one of two things: <ul style="list-style-type: none"> ○ figure out a good GOTV model score threshold to stay above, and generate a SMS universe based on that (limiting to people with high confidence cell phone data on the file) ○ figure out how much budget you have for a SMS GOTV program and generate your SMS universe based on those parameters (working down from highest score down until you hit the target number of voters will high-confidence cell phones on file)
direct mail	<ul style="list-style-type: none"> ● with a GOTV model, and depending on how it's delivered (typically, voters are scored with an estimate of the number of net new votes they could produce if contacted), you can do one of two things: <ul style="list-style-type: none"> ○ figure out a good GOTV model score threshold to stay above, and generate a direct mail universe based on that ○ figure out how much budget you have for a direct mail GOTV program and generate your direct mail universe based on those parameters (working down from highest household score down until you hit the target number of mailboxes) ● if you have access to a mail readership predictive model, you can further narrow, or target, your direct mail universe to only include mailboxes that are more likely to actually read their mail
digital ads	<ul style="list-style-type: none"> ● with a GOTV model, and depending on how it's delivered (typically, voters are scored with an estimate of the number of net new votes they could produce if contacted), you can do one of two things: <ul style="list-style-type: none"> ○ figure out a good GOTV model score threshold to stay above, and generate a digital voter contact universe based on that ○ figure out how much budget you have for a digital GOTV buy and generate your digital GOTV universe based on those parameters (working down from highest GOTV score down until you hit the target number of people) ● if you have access to a digital consumption predictive model, you can further narrow, or target, your universe to only include people likely to consume digital media regularly