How to Build a Strong Predictive Model in 5 Steps

Go from unstructured data to reliable models with these 5 steps.

rapid insight®

Contents

At Rapid Insight, we train and assist thousands of our users each year with their data prep and predictive modeling projects.

We support their work by providing intuitive software, offering feedback on their models, and contributing tips and techniques for improvements in process and design.

Over years of close work with our users, we've observed and documented five steps that lead to the best predictive modeling outcomes.

Here are the 5 Steps to Building Strong Predictive Models.

Want to learn more about predictive modeling with Rapid Insight? Visit us on the web!

rapidinsight.com/solutions/predictive-analytics/

STEP 1:

Define Your Objective

What's the model's purpose and what do you hope to achieve with it?

Step 1: Define Your Objective

In this critical first step, consider the model's purpose and what you hope to achieve with it, then clearly define and record that information.

Follow these steps before building your model to ensure it is built with intention from the ground up:

- **Decide on your Y-Variable**
- **Define the Purpose of your Model**
- **Decide who should be involved in the discussion**

The **Y-variable** is the outcome you intend to predict with your model.

Step 1: Define Your Objective

Decide on your Y-Variable

There are two key considerations when determining your Y-variable. Consider the following:

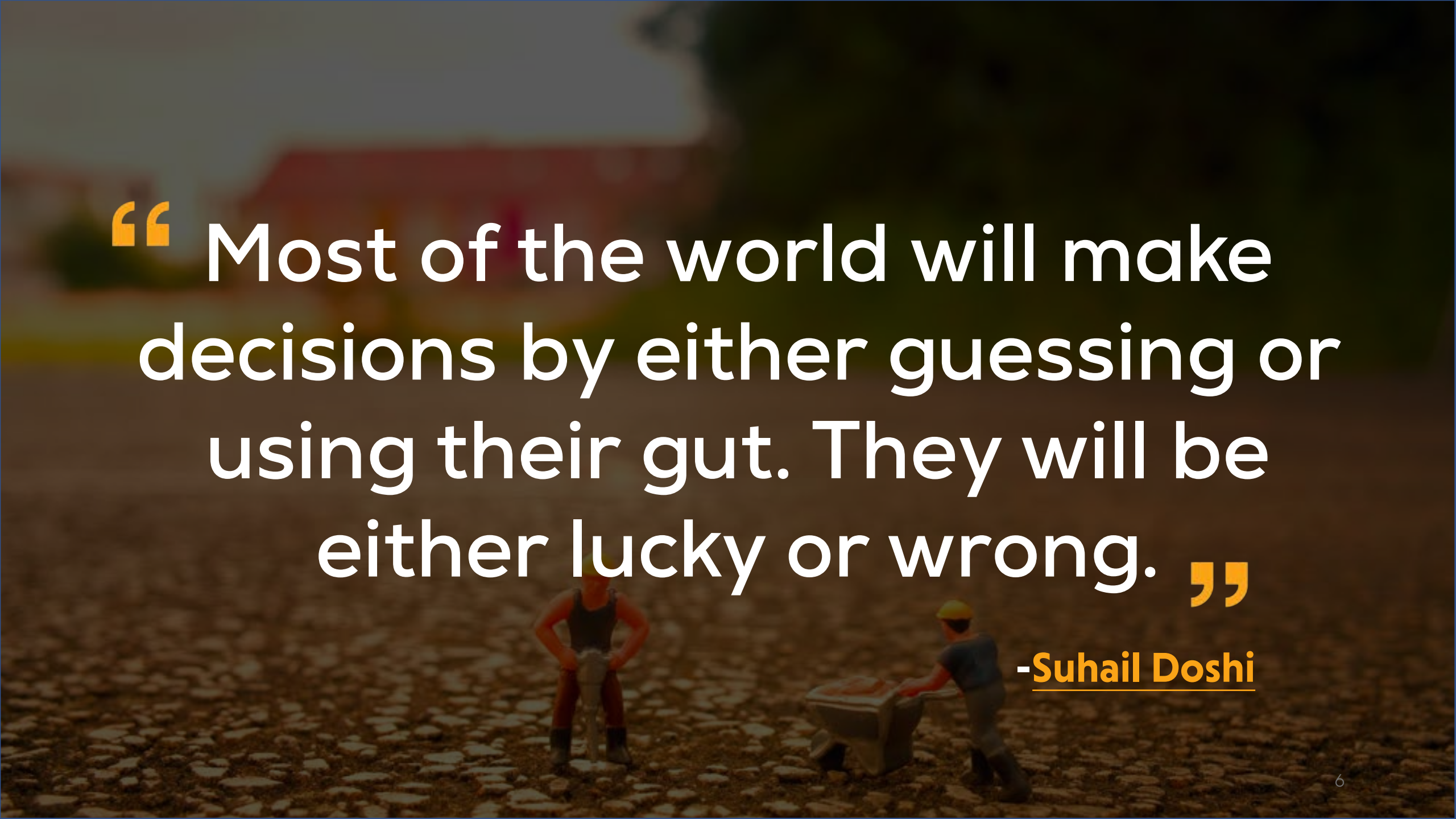
→ **What outcome do you want to predict?**

→ **At what point in time will you apply your prediction?**

Answer these questions to build a foundation for future decisions about your model.

Identifying the **point in time** when you want to apply your prediction will determine what data you require to make an accurate prediction.

For example: predict Fall customer churn using historical Fall data.


“ Most of the world will make decisions by either guessing or using their gut. They will be either lucky or wrong. ”
-Suhail Doshi

Step 1: Define Your Objective

Define the Purpose of your Model

This fundamental step serves as a basis for all future decisions. Consider the following:

- **How will your organization implement this model?**
- **How will findings change the way you do business?**
- **Do you plan to use the model once or on an ongoing basis?**
- **How will users access the model?**

Step 1: Define Your Objective

Decide who should be involved in the discussion

Decide who at your organization falls into each of the following categories:
Note: Some organizations may find that one person fills multiple roles.

- → **Model planners**
- → **Model builders**
- → **End users**
- → **Data governors**
- → **Beneficiaries and stakeholders who do not directly use the model**

Then, decide when each party needs to be involved in the process.
It may be best to convene the full group for a preliminary meeting to ensure the model will achieve its goals.

STEP 2:

Establish Data Access

Identify, locate, and access the data you will need for your project.

Step 2: Establish Data Access

With a clear objective in mind and the right stakeholders identified, the next step is to identify, locate, and access the data you need for your model.

You may run into delays or problems if you do not establish a list of necessary data (and a plan to access it).

To assemble your list:

- → **Identify the data you have and the data you need**
- → **Determine where your organization stores data**
- → **Make a plan to access necessary data**

According to Techopedia, good data access implies that:

"Users who have data access can store, retrieve, move or manipulate stored data, which can be stored on a wide range of hard drives and external devices."

Identify the data you have and the data you need

Put structure to the process of gaining access to data. Develop a list of the following categories of data:

- **Data that is required and currently accessible**
- **Data that is required but *not* accessible**
- **Data that would be nice to have, but isn't critical**

Next, arrange the list in order of urgency. Which items on the list are the most critical to your model?

Step 2: Establish Data Access

Determine where your organization stores data

At most organizations, data lives in many disparate systems and storage locations.

For a successful predictive model, you need to incorporate all data into a single modeling dataset.

Take these steps:

→ Create a list of all systems your organization uses that may contain information related to the outcome you plan to predict

→ Match this list up with your list of data needs to identify the systems and files you'll need to connect to

HELPFUL RESOURCE

→ Learn to Blend Database Systems to Prepare for Predictive Modeling

Step 2: Establish Data Access

Make a plan to access necessary data

Whether your required data is stored in a cloud-based application, on a server, or on local drives, it's time to incorporate it into a single dataset.

Consider these actions:

- **Identify and connect with data owners or system administrators. Explain your project and the reason for your data request, then establish credentials or a system for sharing the desired data**
- **Direct access is your best option, but that isn't always available. You may need to find workarounds, such as periodic uploads or extracts**
- **Start by working on your most urgent needs (data without which your model cannot function)**
- **If needed, bring in stakeholders with greater leverage to assist you**

STEP 3:

Prepare the Data

To build a trustworthy model, the data going into it must be clean and well-structured.

Step 3: Prepare the Data

To build a trustworthy model, you must ensure that the data going into it is trustworthy (which means clean and well-structured).

Data prep is often the most time-consuming, dreaded part of the model building process, but modern tools, such as Rapid Insight's Construct, make the process easy and intuitive.

There are three important questions to answer when prepping data for a predictive model:

1. **Ensure that the data is reliable**
2. **Decide if the model will be used once or repeatedly**
3. **Document the data prep process**

RELATED WHITEPAPER

→ Data Quality & the Path to Stronger Decision-Making

Better data leads to better decision-making, and it all begins with understanding how data is defined, measured, and prepared for effective reporting & modeling.

Ensure that the data is reliable

Data preparation helps you verify that your data is trustworthy. You must be able to defend why you included certain data in the model. Ensure that it was properly collected, or decide how to proceed if it was not.

Be sure to ask:

- **Is there an audit trail available?**
- **Was the data uniformly collected across the whole population?**
- **Were retroactive changes made to the data during an update or audit? If so, does this impact the data's reliability?**

No matter how useful data is to a model, if it's unreliable, it's best not to include it.

Decide if the model will be used once or repeatedly

Do you plan on updating and tweaking the model as time goes on?

If so, it is crucial to make your data preparation process repeatable.

This means that when you update this model or drop new data into it (next month or next year), you are able to quickly replicate (or ideally, reuse) the established process.

→ **Modern software, such as Rapid Insight's Construct, automatically saves your data prep process for future use**

→ **In lieu of such software, ensure you record your steps and process to follow again in the future**

RELATED POSTS

→ 6 Top Data Wrangling Tips

→ How An ETL Tool Can Help Your Organization

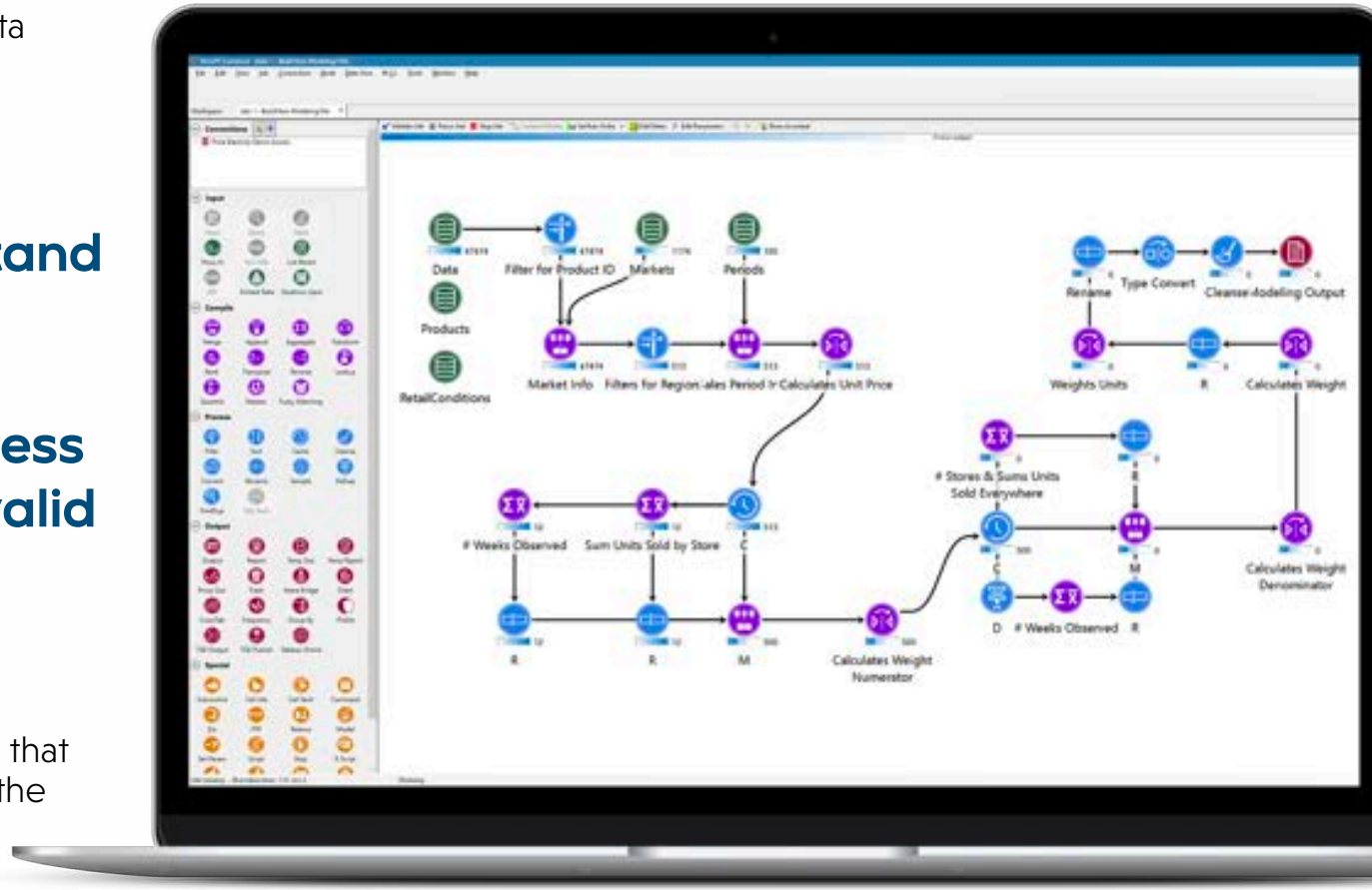
→ Choosing the Right Data Preparation Software for Your Organization

Document the data prep process

You should be able to explain the origins of your data and your data prep process. Ensure that you document both.

- **Create a data dictionary so that end users and future analysts can understand your dataset**
- **Build a diagram or flowchart of a process you intend to repeat. This will ensure valid procedure and help you communicate the process to stakeholders**

Rapid Insight's Construct features a visual, drag-and-drop interface that is easy to use, both when prepping the data and when explaining the process to stakeholders.



construct

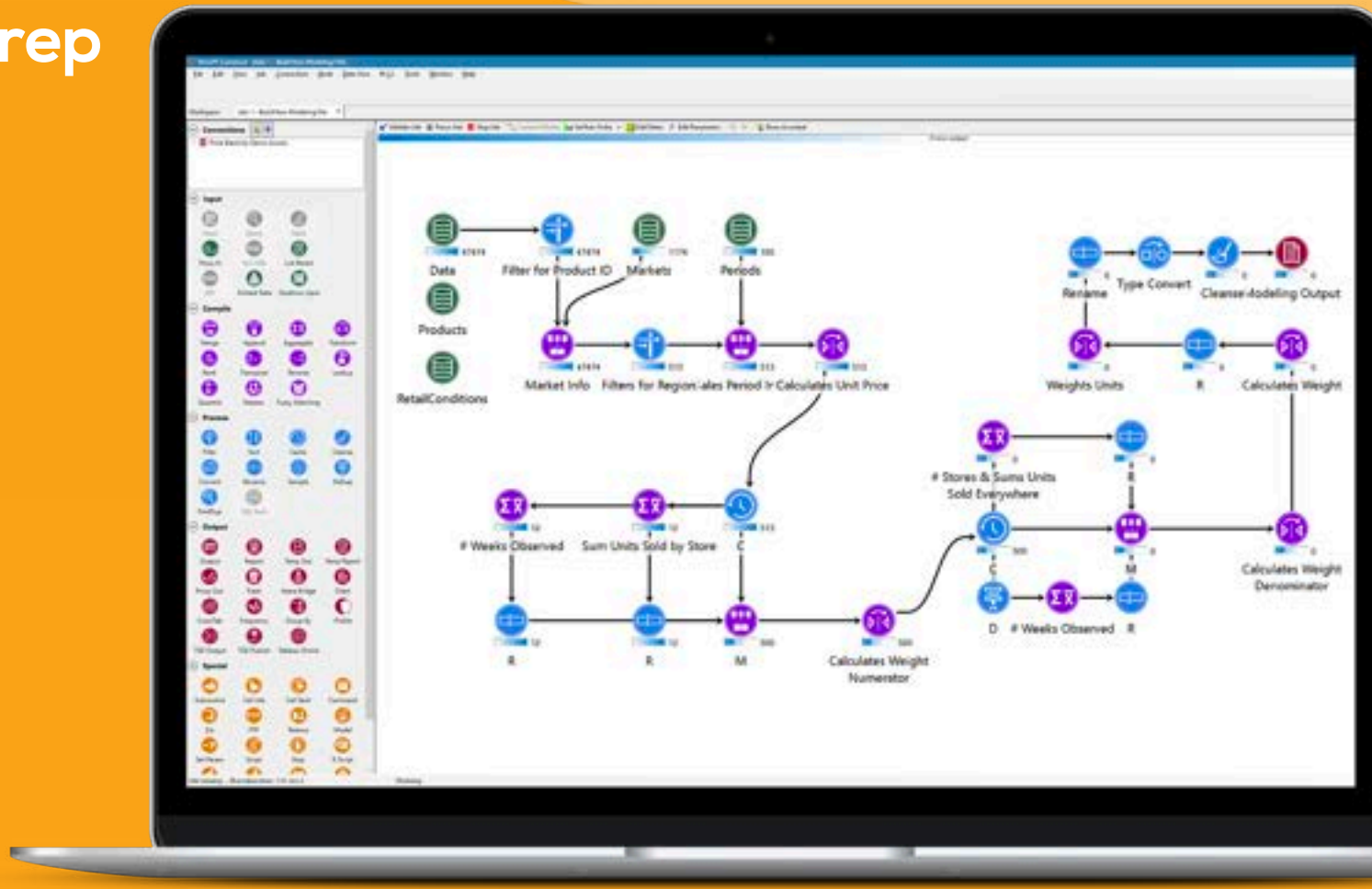
A Data Prep Tool for **Everyone**

Good models rely on **data prep**

With Rapid Insight's Construct, seamlessly connect to any data source, then easily develop structured, trustworthy modeling data.

Construct is powerful enough for any task and intuitive enough for any user. Its drag-and-drop, node-based interface makes combining, preparing, and exploring data a transparent, repeatable, enjoyable process.

Learn More



STEP 4:

Build the Model

It's time to build the model and establish an understanding of what your organization should expect from it.

Step 4: Build the Model

With clean data in hand, it's time to build your model and establish an understanding of what the organization should expect it to deliver.

To build the model itself, we advise using a modern predictive modeling software that makes the process easy, such as Rapid Insight's Predict. An automated modeling software opens up the process to wider range of subject matter experts and makes the process fast and repeatable, allowing you to build many models for critical projects.

However, even Microsoft Excel (paired with a solid knowledge of statistics) is capable of building models.

When building and tuning your predictive model, follow these steps:

1. **Identify the assumptions you and your team have about the outcome**

2. **Prepare the model for implementation**

RELATED POSTS

→ How to Interpret a Decile Analysis

→ Data on the Edge: Handling Outliers


“ Data beats emotions. ”
-Sean Rad

Identify the assumptions you and your team have about the outcome

Before deploying your model, it is important to identify assumptions.

Often, the model will validate these assumptions. However, it is critical to document them to see if the data supports or challenges them.

It is especially important to note assumptions which currently inform business decisions, or which might have triggered the modeling effort to begin with.

Predictive data can be used to depersonalize conversations about such business decisions: if the data doesn't support an assumption, the case to change course is stronger (and less personal).

RELATED POST

→ 9 Ways to Prevent Data Bias in Predictive Models

Prepare the model for implementation

Once a model is built, it's time to implement it so that it can start driving decisions.

Plan implementation with end-users in mind. Establish how, when, and where users will access results.

→ Establish where end-users will view predictions

Identify where users will go to access the model's results, whether that's inside of a database, as a standalone list, or in another location. This will tell you where scored records need to end up and what format to store them in.

→ Streamline data prep for scoring

To ensure new records become scored records as quickly as possible, establish a repeatable, automated process for data-cleaning and new-variable generation.

→ Automate your model's calculations

Build a schedule, a macro, or a derived field to automate future calculations. Whatever option you choose, plan for it to be automatic or routine. If you must manually click a button, you're keeping users waiting.

predict

Embrace the **Future**

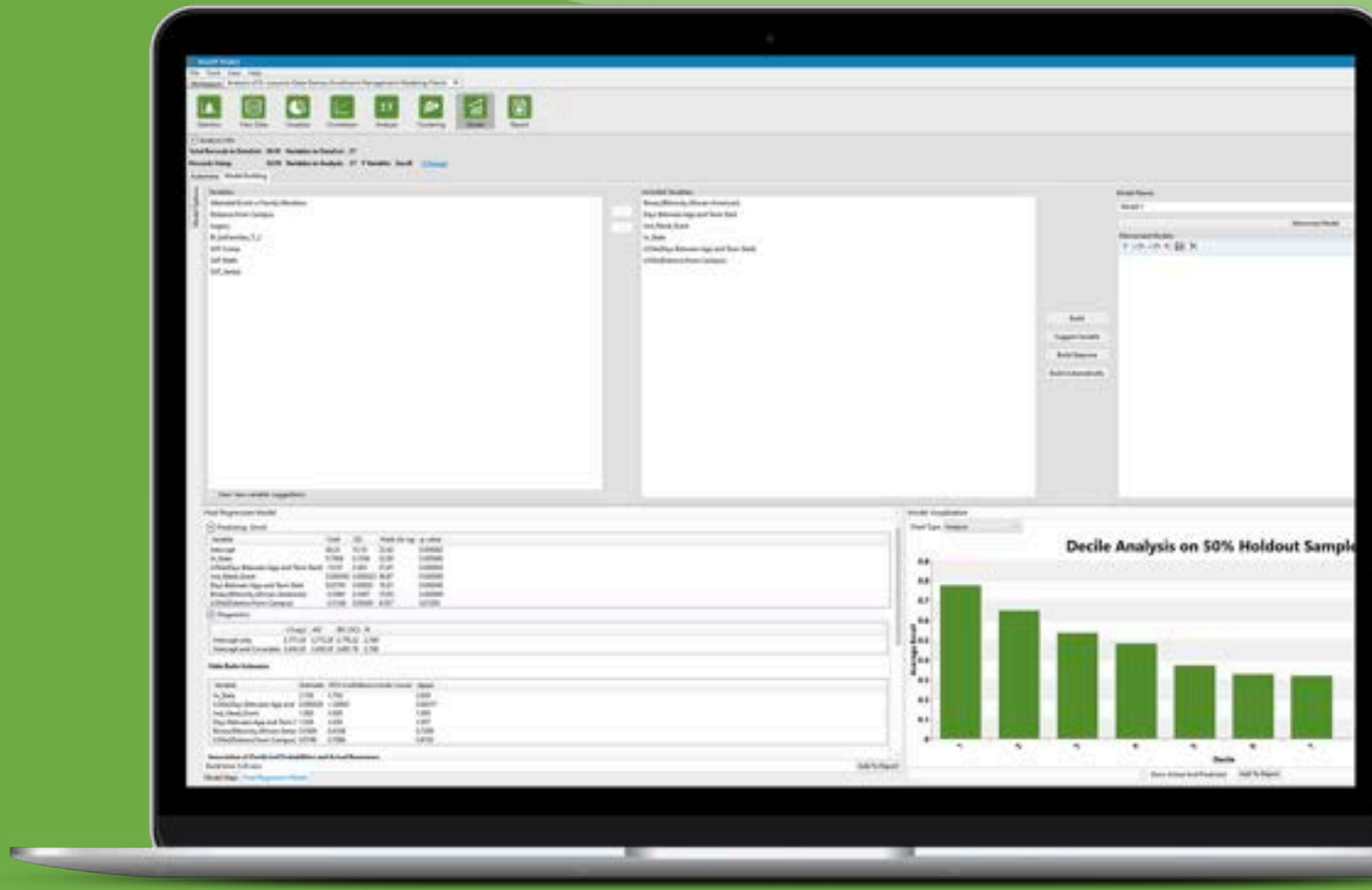
With automated **Predictive Analytics**

Rapid Insight's Predict makes building predictive models easy.

With a single click, Predict automatically mines your data, analyzes the depth of relationships among variables, and creates the strongest predictive model possible.

Predict enables users of any background to build models for every application. See the inner workings of the model. Make adjustments based on stakeholder feedback. Deliver the results your organization needs to make data-informed decisions.

Learn More



STEP 5:

Utilize the Model

Put your model to work and get results in the hands of end users.

Step 5: Utilize the Model

It's time to put your model to work and get results in the hands of end users.

Consider how to most effectively present the model's predictions to decision-makers at your organization.

These are the important questions to consider at this stage:

- **Implement the model**
- **Decide how best to present or distribute the model's results**

Implement the model

As you roll out the results of the model, consider how your organization will use its predictions to inform decisions. When making these decisions, consider the resources you have available and what volume you can handle.

- **Decide how precise or general the results should be for your given audience**
- **If needed, create "bins" and focus on certain score ranges**
- **Decide if end users need to see probability scores**
- **Decide who should have access to the data (and who should not)**
  - **Evaluate this question with both data security and privacy in mind**

Decide how best to present or distribute the model's results

What is the best format to deliver the model's results to your stakeholders?

There are a range of options, including:

- → **Ad hoc reports**
- → **Scheduled reports**
- → **Formal presentations**
- → **Simplified dashboards**
- → **Stylized visualizations**

Modern data-sharing platforms, such as Rapid Insight's Bridge, make custom dashboards accessible to designated users at any time, on any device.

A **data-informed** culture

bridge

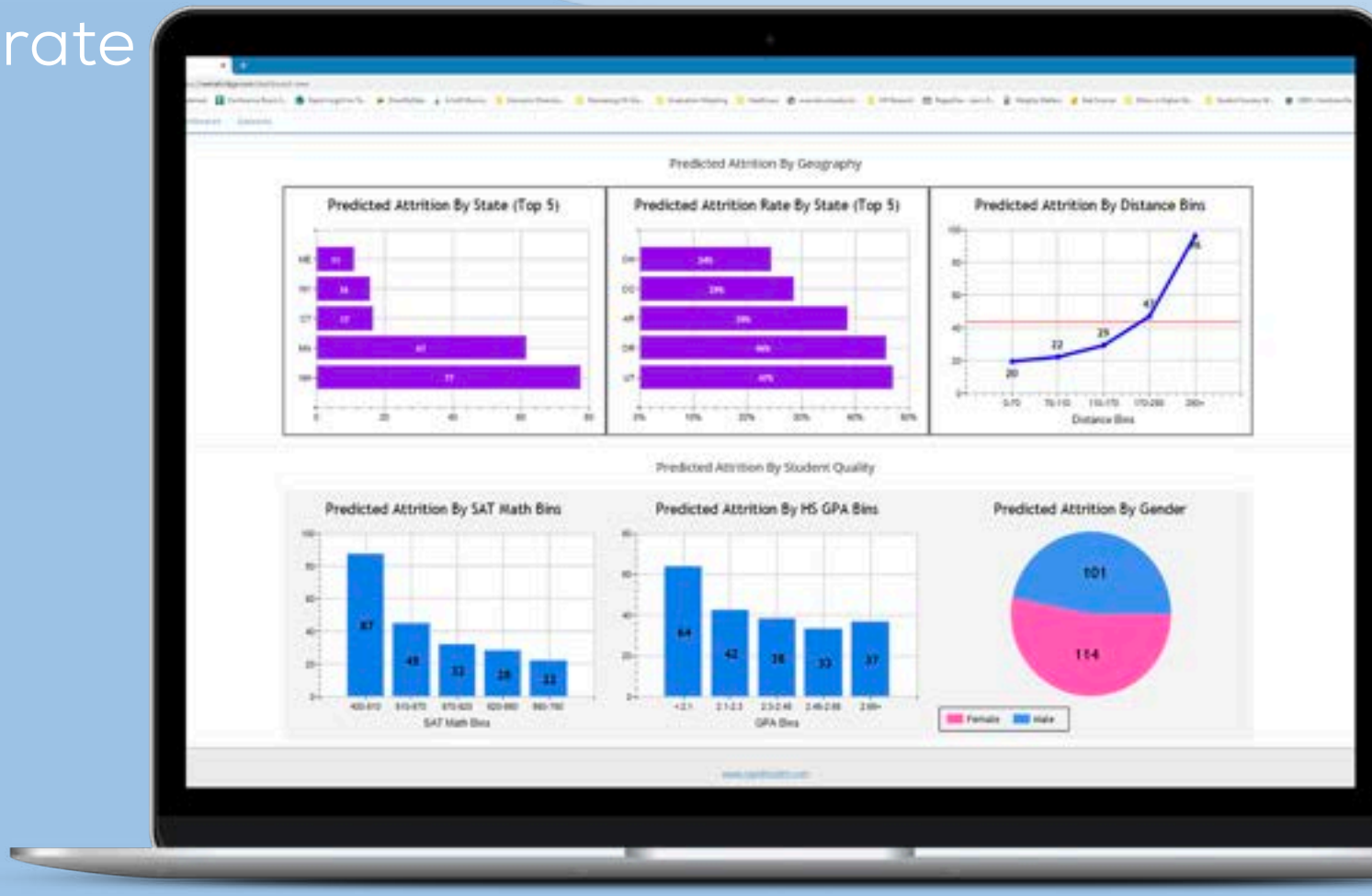
Visualize, interpret, collaborate

Rapid Insight's Bridge features a clean, user-friendly interface that enables quick data visualization through attractive auto-generated reports and custom dashboards.

Enable everyone in your organization to access the information they need in the format they want to see it.

Democratize the results of your predictive modeling projects.

Learn More



Predictive Modeling Checklist

Print this page... and don't forget to refer back to each chapter for guidance on how to thoroughly complete each step!

☐ 1. Define Your Objective

- → Decide on your Y-variable
- → Define the purpose of the model
- → Decide who should be involved in the discussion

☐ 2. Establish data access

- → Identify the data you have and the data you need
- → Determine where your organization stores data
- → Make a plan to access necessary data

☐ 3. Prepare the data

- → Ensure that the data is reliable
- → Decide if the model will be used once or repeatedly
- → Document the data prep process

☐ 4. Build the model

- → Identify the assumptions you and your team have about the outcome
- → Prepare the model for implementation

☐ 5. Utilize the model

- → Implement the model
- → Decide how best to present or distribute the model's results