# “High Performer Modeling” Module Transcript

## Chapter 1

### Intro, Topics Covered, & Learning Outcomes

Hi. I'm Hayley, and I'm part of the One AI team here at One Model. In our previous modules, we've covered the basics of machine learning and the best practices for building models. Now we're excited to bring it all together and create a high performer likelihood model using the One AI recipe. This model is perfect for anyone looking to predict who will be a high performer and understand the attributes that drive these behaviors.

We will cover an overview of the high performer machine learning model, some key considerations before beginning the model-building process, step-by-step instructions for building a high performer likelihood model in One AI with a recipe, and we will close with a review of some of the important insights that can be drawn from this type of model in One Model.

After completing this module, you will understand and address important considerations before beginning the model-building process ensuring effective and accurate predictions. You will be able to confidently create a high performer likelihood model that is relevant to your organization using the One AI recipe, and you will utilize the insights drawn from the model to identify and develop high potential employees, optimize performance management strategies, and enhance overall productivity.

## Chapter 2

### Overview & Purpose

Section 2 - Overview and Purpose

The high performer model is designed to help you predict the likelihood of whether an employee will be a high performer during a specified amount of time in the future, for example, within the next year. By analyzing attributes such as employee demographics, job satisfaction, engagement, previous performance metrics, manager’s performance metrics and team performance metrics this binary classification model identifies individuals with high likelihood of being high performers.

The two outcomes that the model will classify employees into are high performers and not high performers or average-to-low performers. The instances are the individual employees usually identified by their unique person IDs from the employee table.

The model creator selects the model population, the prediction period, and the input features used to train the model to make its predictions. Input features include core attributes and individual and team generative attributes.

Additionally, model results will include the top drivers of high performers and average-to-low performers, which allows you to gain deeper insights into the underlying factors that influence employee performance. This enables organizations to design more effective management practices and foster a work environment that nurtures talent and maximizes potential.

Some key use cases for the high performer model include talent identification and development, to include succession planning to identify potential leaders for appropriate development as well as training and development to ensure high potential employees are nourished to maximize their growth and lower performing employees are helped to improve.

Performance management - leaders can use predictions to inform performance reviews and provide targeted feedback, as well as setting personalized performance goals for employees based on their likelihood of becoming high performers.

Insights from these models can also be applied in recruiting and hiring because once we know what drives a high performer, recruitment efforts can focus on candidates with similar attributes. Those implementing retention strategies can use insights from these models to implement targeted retention programs for high potential employees to reduce turnover.

And finally, these insights can be considered in the team composition and dynamics realm. Leaders can use model results to assemble teams with a mix of high performers to ensure balanced and effective team dynamics.

## Chapter 3

## Key Considerations Before Model Building

Section 3 - Key Considerations Before Beginning the Model-Building Process

Before creating a high performer model, there are some important factors that should be carefully considered to ensure its effectiveness.

First, define your objectives and goals for the model. Determine if the model aims to confirm hypotheses, conduct exploratory analysis, or replicate a previously built model. This will guide your approach. For example, exploratory models might use various input features and historical data lengths, while hypothesis-confirming models will use specific historical data periods and features believed to drive performance.

Next, you'll want to ensure you have accessible and reliable data to train the model. Verify that the necessary data sources are integrated and validated in One Model before beginning. Consider if there is sufficient historical data on performance and performance management to make accurate predictions. For example, if your organization has not conducted performance reviews or has not brought this data into One Model, the model will not be able to learn what makes an employee a high performer.

Next, consider your organization's performance cycles. While one year is a common prediction period, you should consider your performance review cycle and timing for when employees of interest are going to be reviewed in the future. This is also important to consider when you determine how much historical data the model will need to learn effectively. Typically, 1-2 years is optimal, but this can vary based on how often you perform performance reviews.

And finally, ensure you have the resources to build and maintain the model. While One AI simplifies the model building and visualization process, ongoing maintenance is essential. This includes rerunning and refining the model when new data becomes available, creating effective visualizations and storyboards, incorporating feedback from stakeholders, and monitoring performance. This maintenance effort will vary based on whether the model is for long term use or a one-time analysis.

Next, we will move over to One Model for a demo on building a high performer model in One AI with the recipe.

## Chapter 4

## How to Build in One AI

Section 4 - How to Build a High Performer Model in One AI

Before getting started, you will need the CanAccessOneAIMenu application access role to be able to build models in One AI. If you can see the One AI tab in the main ribbon menu, you should be all set. If not, get with your site administrator to get this permission.

Once you're ready to start building, click 'One AI' in the main ribbon menu. This is where you can edit or create models. If your organization has previously created any models, they will be listed here. To add a new model, click 'Add Machine Learning Model' in the upper right hand corner.

First, give your model a unique name in the Display Name field. Once the model has been created and saved, the display name cannot be changed, so choose wisely.

Next, click 'One AI Recipe' and then 'Configure One AI Recipe'. This brings you to the One AI Query Builder screen. Under "What are you interested in predicting?", select 'High Performer' from the list of available recipes. Once selected, the problem statement is populated with the information we need to provide to build the model. One AI simplifies most settings by automatically selecting defaults, so most of the configuration focuses on data framing.

In step one, we select the column or metric to define high performers. This should be a performance review occurrence metric filtered to however your organization defines high performers. We don't want to use a headcount metric here because step one is all about defining the event One AI will predict. First, I will pick ‘Metric’ from the first dropdown because columns should only be used if they were created by your data engineer specifically for One AI purposes. Then I am going to select the Performance Review Occurrences metric. This metric includes all performance rating occurrences and therefore all levels of performance ratings, but we want to predict only on high performers; so, I will filter it with the performance rating dimension to only include high performance ratings.

If you do not have a performance occurrence metric, you can build one by using the occurrence column from the performance review table. Something like this works well. If you do not have this column, chat with your CS team to get it added.

Next, we will select the headcount population, which is the group of employees for which One AI will predict high performer likelihood. You can use a general headcount end of period metric and filter down or a specific headcount metric like managers or engineers to get to your group of interest. I'm going to use a general headcount end of period metric and filter down when I'm ready.

Next, choose a unique identifier for each instance in the model. This is typically person ID from the employee table, but you can use any unique ID that you'd like, including worker ID or a WID or any other unique ID. One AI will verify the ID's uniqueness for the model. Next, we will select the population date. The population date anchors the query with the predict frame and the train test dataset offset from it. Common choices are today, the end of last month, or a static date. You should avoid dates too far in the past since we already know the outcomes. I am selecting today, which means that this model will use history going back starting from today and will predict a year in the future also starting from today.

And finally, use filters to refine your headcount population. Instead of running the model on the entire organization, you should segment it by factors like department, country, or job level. This is best practice because motivations and drivers may vary drastically by group. For example, I'm going to use the country filter, and I'm going to filter it to only include employees in the United States.

Next, select how far into the future you want to predict from the population date. One year is the most common choice because it's helpful to know who will be a high performer relatively soon, perhaps in the next performance cycle, but you can adjust it to fit your needs.

Then choose how much historical data to use for model training. The previous step also determines the increments for how far back in time we will go to train the model. If you selected 1 year, then training intervals are in 1 year increments, meaning 1 equals 1 year, 2 equals 2 years, etcetera. I usually start with 1 year and increase to 2 or 3 years to see if it improves performance. Typically, less is more as training on all available data may not be beneficial due to changing organizational dynamics and employee behavior.

Adding meaningful labels to the values you are predicting is optional but will simplify reading your Results Summary and storyboards greatly. In this model, null values or 0s indicate not high performers, while values greater than 0 indicate high performers. Click 'Load Target Metric Values' to set overrides. You can name the groups as you like, such as "High Performer" for 0s, and "Average-to-Low Performers" for greater than 0, but just make sure there are only two unique labels total. Without this step, predictions will be sorted into 1s and 0s, making results and storyboards difficult to interpret for end users who don't know which label corresponds to which value.

In the next step, we select the core attributes, which are the input variables that form the dataset that the model learns from and uses to make predictions. We want to aim for balance here. Enough attributes for robust predictions, but not so many that the results become difficult to interpret.

There are four scopes available. First, we have 'None' where only the unique identifier is included automatically, and then you must manually add other attributes. Next, we have 'Narrow', which includes all attribute columns from the table where the unique identifier originated from, which is the employee table in this case. Then we have 'Balanced', which includes all attribute columns within one join from the table where the unique identifier originated, which again is the employee table. And finally, we have 'Broad', which includes all attribute columns within a possible join path to the table the unique identifier originated from.

If none of these settings fully meet your needs, you can manually include or exclude columns using the checkmark button to include additional columns in the Excluded section, or the X button to exclude columns from the Included section. You can use the search bar if you know what the columns are called to find them quickly.

For this model, I recommend starting with a narrow scope to include the full employee table, but manually adding in all performance tables' columns, manager's performance tables' columns, the event tables' columns, and any other columns that you are particularly interested in. Once you are comfortable interpreting the model results, you can move into a balanced scope if desired. Broad tends to include attributes that may be harder to relate to the model as they get further away from the employee table. If you find that a table you're looking for is unavailable, it means using it would cause duplication or there's no join path. The reason is explained in the "Reason Column". In some cases, you can work with your CS lead and data engineer to make unavailable tables available.

In the next step, you can select generative attributes you want to use in your prediction. Use the checkbox to the left of each attribute to include it as an input variable. This step is optional, but highly recommended because generative attributes often enhance the model performance and robustness. You can also create, edit, and delete generative attributes in this step. For more information on how they work and how to build them, check out our module on generative attributes.

In the next step, click 'Generate Data Statistics' to verify that all your previous selections are valid. This helps identify errors that cause the model to error while running. You will also receive valuable data exploration information about the train and test dataset, such as row count, standard deviations, and min and max values. This helps identify columns you may want to exclude from your dataset.

As you can see, we received the green light "Success" so we can proceed. If One AI finds a problem, it will display "Action Needed" in red with details about the errors that need to be corrected before you can run your model.

In the final step, you can download the Train/Test Dataset, Predict Dataset, or both that were generated by this recipe. This is useful for analyzing the data in external tools or for investigating errors encountered in the verification step.

Then you should review your selections from each step. Once it looks good, click the save icon in the upper right hand corner to return to the machine learning model screen. Here you can manually configure model settings. Refer to our modules on advanced configuration and global settings if interested. At minimum, I highly recommend generating SHAP from the Global Settings in order to visualize this data on storyboards.

Once complete, scroll to the bottom and click 'Create' here. If you don't click create and navigate elsewhere in One Model, you will lose your work from the One AI Query Builder screen.

Congratulations! You can now create high performer models in One AI. The next step is to run your model and then evaluate your results, refine your model based on those results, and then deploy and share the insights.

## Chapter 5

## Insights Drawn

Section 5 - Insights Drawn From a High Performer Model

High performer likelihood models provide several insights that can help your organization make good decisions.

First and foremost, you will receive data on the model's predictions made at the instance level. These can and should be aggregated to larger groups for better accuracy. Without creating any visualizations, you can view individual predictions by dataset ID in the Results Explorer.

You can also use storyboards to analyze individual employees or groupings within the model population examining which features drive behaviors, how it compares to the mean, and an explanation of these drivers. For example, this chart here is filtered down to the finance orgs population.

You can also analyze correlation data from the EDA report to see how your input features interact. This can help you understand how your data is related to each other and uncover interesting patterns to provide even more insights on what makes a high performer perform so well.

You can also understand and visualize the factors influencing employees to either achieve high performance or not. If SHAP is enabled in your model and incorporated into storyboards, you can view the top drivers for both classes in order of importance.

Additionally, you can analyze likelihood by groupings within your model population. Any dimension that can be used to filter the headcount metric can break out your model population to identify likelihoods for high, medium, and low performance by grouping. For example, using the job level dimension, you can see that intermediate staff have the largest population of employees predicted to be low performers.

You can also analyze individuals within your model population to get targeted insights about specific employees. You can view the predicted likelihood of high performance, performance rating history, feature impacts, and feature value differences from the mean.

And finally, you can analyze by-name lists of folks predicted to be high performers, medium performers, and/or low performers, and view other information about these employees, like who their manager is, what org unit they belong to, their performance history, and much more.

These are just a few examples of insights that can be drawn from high performer models. Many more are available depending on your objectives and what questions you are using the model to try to answer.

## Chapter 6

## Conclusion & Thanks

We've explored how to build high performer likelihood models using One AI. By understanding the key considerations, common use cases, and the powerful insights that can be drawn from the model's results, you can confidently create and deploy high performer models for your organization. These models will help you identify and develop high potential employees, optimize performance management strategies, and gain valuable insights to enhance overall productivity. Happy modeling!