# Salesforce Einstein Model Cards

Salesforce, Spring ‘24


Last updated: April 5, 2024

# CONTENTS

# MODEL CARDS OVERVIEW

Model cards help ensure transparency around Salesforce AI. Each model card gives details about the performance characteristics of a trained machine learning (ML) and artificial intelligence (AI) model. For example, learn about inputs, outputs, the conditions under which the model works best, and ethical considerations in its use. By providing this type of transparency, we aim to help developers, customers, journalists and industry analysts, policy makers, advocacy groups, and consumers better understand the impact of our AI on individuals, communities, and society.

For more information, including a list of available Salesforce model cards, see Model Cards for AI Model Transparency on the Salesforce Einstein blog.

Note: This information is subject to change. If you download this guide, you can always get to the most current version at *https://resources.docs.salesforce.com/latest/latest/en-us/sfdc/pdf/salesforce_ai_model_cards.pdf*.

## Einstein Opportunity Scoring

Einstein Opportunity Scoring helps sales teams and managers focus on the right opportunities so they can close more deals. Each opportunity is given a score from 1 to 99, which indicates how likely a deal is to close. Scores are available on opportunity records and list views. Customers can use scores with reports, Process Builder, and workflows. If you use Collaborative Forecasts, opportunity scores are also available on the forecasts page.

For more information about how Einstein creates opportunity scores, see Salesforce Help.

If you have questions or comments about the Einstein Opportunity Scoring model, contact Salesforce Customer Support or reach out to the Trailblazer community.

EDITIONS

Available in: Lightning Experience and Salesforce Classic.

Available with Sales Cloud Einstein, which is available in **Performance** and **Unlimited** Editions, and for an extra cost in **Enterprise** Edition

Available to eligible customers for no extra cost in: **Enterprise**, **Performance**, and **Unlimited** Editions

## Model Details

**Team Developing the Model**

Sales Cloud Einstein engineering

**Single-Customer and Global Models**

Single-customer models are built using only the specific customer's data. Customers must meet the data requirements in order to use a single-customer model. If they don't have enough data, the global model is used by default.

The global model is built using data from multiple eligible Salesforce customers and is anonymized. Eligibility is determined by a set of factors, including having sufficient data. Salesforce customers can opt out of having their data used by the global model. Customers who opt out of contributing their data can still use global models.

Both the single-customer and global models work the same way. Both are retrained on a regular cadence. The model card information applies to single-customer models and global model, unless otherwise noted. For additional information, see Training and Evaluation on page 3.

**Initial Release and Updates**

- Single-customer model: Spring '17
- Global model: Spring '21
- Minor changes can occur throughout the release.

- Major changes can occur and are communicated in the Salesforce release notes.

**Model Type**

Binary Classification

# Intended Use

**Primary Intended Users**

- Sales managers and sales reps
- Sales operations

**Primary Intended Uses**

- Identifying at-risk opportunities and revealing neglected deals that can help sales reps hit their quotas
- Prioritizing opportunities when there's a high volume of deals
- Understanding sales patterns across all opportunities

**Out-of-Scope Uses**

Opportunity Scoring is meant to be used to manage deals and deal flow. Uses other than managing deals are out of scope. Also, nonstandard sales processes or nonstandard use of the opportunity Stage field are out of scope. Out-of-scope examples include users who create records for deals only as the deals are being won and users not applying the Stage field accurately.

# Factors

When building the model, we look at past closed opportunities (both closed-won and closed-lost). More specifically, we look at the following data. Be aware that some of the data is used for only single-customer models.

- Each opportunity's record details (both standard and custom fields), history, and related activities
- The related account's record details and some record history
- Details about the related products, quotes, and price books

When an individual opportunity receives a score, the score is shown with several key factors. The key factors are relevant to the specific opportunity and indicate the reasons for the particular score. Providing the reasons for the score helps users makes sense of the score and take any necessary action.

Sometimes an opportunity has a score but doesn't show any or all key factors. This result can occur when there are too many minor factors for any of them to help users understand how the score was calculated. Or, it could occur when the factors are complex and, therefore, too difficult to summarize.

Factors contribute to the score in both positive and negative ways. A factor that is positive for one company might be negative for another. For example, a deal with an enterprise customer requires extra steps, such as legal review and a longer negotiation process. These extra steps delay the close date. But, because the delayed close date signals that the deal is progressing, the delay is associated with a higher score. However, for a deal with a small- or medium-sized business (SMB) customer that doesn't include extra steps, a delayed close date is associated with a lower score.

# Model Performance Metrics

To monitor and improve the quality of the model, we gather model performance metrics. Metrics include the correlation of output scores to historical observations. All metrics are aggregated and anonymized. Customers are responsible for monitoring the accuracy of the opportunity scores.

The main model performance metric is Area Under Precision-Recall (AUPR) curve, which optimizes for precision and recall over different sets of scoring thresholds.

Also, to prevent a skew towards a narrow range of scores, we use a standard deviation (STD), which calculates the score spread.

To improve model accuracy and the overall customer experience, we sometimes change the model performance metrics.

Examples of model performance metrics include:

- Lift Top Bucket: Indicates uplift in win rate for the opportunities within the 80th percentile of scores. An uplift in win rate is computed versus the overall win rate.
- Lift Bottom Bucket: Indicates negative uplift in win rate for the opportunities within the 20th percentile of scores. An uplift in win rate is computed versus the overall win rate.
- Model Skew: Ensures diversity among scores and that scores aren't consolidated around a narrow range of scores. For example, a high percentage of opportunities shouldn't receive the same score range.
- Key Factor Coverage: Represents the number of opportunity scores that don't show key factors. A score doesn't show factors when there are too many minor factors for any of them to help users understand how the score was calculated. Or, this result could occur when the factors are complex and, therefore, too difficult to summarize.
- Key Factors Diversity: Ensures a variety of key factors types, including factors based on opportunity stage, amount, and close date.

# Training and Evaluation

**Model Training**

To train and evaluate the single-customer model, we use a sample of data from the customer's own Salesforce data. Of the total sample data, 80 percent is used for training and 20 percent is used for testing and evaluation. We train multiple single-customer models with different configurations. We use the one with the best model performance metrics, as determined by the model tournament.

To train and evaluate the global model, we use data from multiple Salesforce customers that represent various industries. Of the total data, 75 percent is used for training and 25 percent is used testing and evaluation.

**Model Tournaments**

To ensure that we use the highest quality models, we evaluate each model against our model performance metrics. This evaluation process is known as a model tournament. First, we evaluate the single-customer models. When certain conditions are met, the performance of the "winning" single-customer model is evaluated against the global model. A combination of Area Under Precision-Recall (AUPR) and standard deviation (STD) is used for the tournament. Even though the global model is built and regularly tested for accuracy, the single-customer model often performs better than the global model.

A single-customer model can't be built if the Salesforce customer doesn't have enough data. In that case, the tournament between the single-customer model and the global model isn't run, and the global model is chosen. To see what data is required to build a single-customer model, see Data Requirements for Sales Cloud Einstein.

**Evaluation Versus Runtime**

When models are built, we evaluate performance against a sample of test data. For Einstein Opportunity Scoring, the sample data includes opportunity history for closed-won and closed-lost opportunities. This evaluation is used as a basis for the performance metrics used in the model training and model tournament. When the model is deployed, it's used to push scores to open opportunities. If the open opportunities have characteristics that are very different from the opportunities used for model training, performance metric results can differ in runtime. Models are re-trained regularly so that they're a better fit for recently closed opportunities.

# Ethical Considerations

We attempted to avoid bias and other ethical risks by not including demographic data in the model. To avoid potentially reinforcing any unknown or unidentified biases in the model, human judgment should be applied to understanding and acting upon the model's outputs.