

Tuning Predictive Scoring

 This article supports Audience Studio - Legacy.

To tune predictive scoring, review the Predictive Scores for your master segment. Depending on your Accuracy and AUC ([Area Under the ROC Curve](#)), you might need to adjust:

- your data set
- predictive features

Tuning predictive features is accomplished using the TD Console, Predictive Scoring Predictive Model page.

- [To Determine Accuracy and AUC](#)
- [Interpreting the Accuracy of Prediction](#)
- [Low AUC Troubleshooting](#)
- [AUC Higher than 0.9 Troubleshooting](#)
 - [Possibility of Overfitting](#)

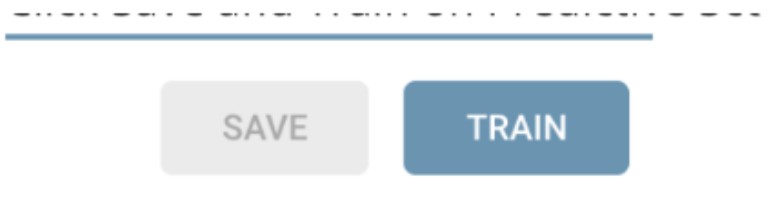
To Determine Accuracy and AUC

1. Define predictive features.



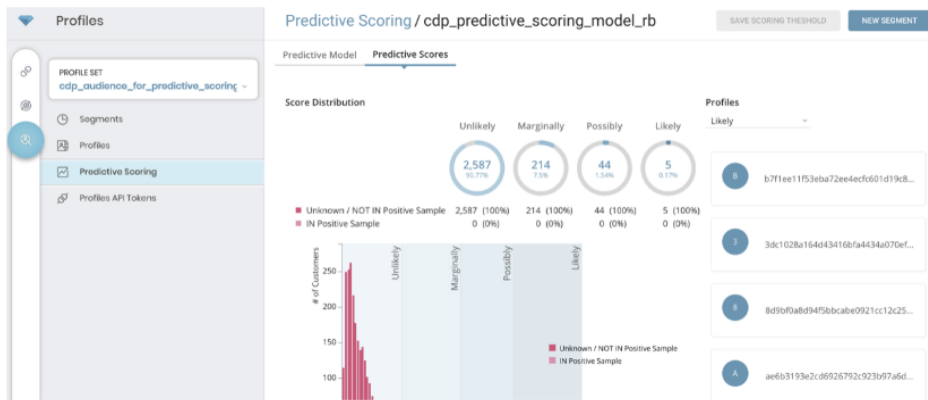
The screenshot shows a user interface for selecting features. It is divided into three sections: Categorical Features, Categorical Array Features, and Quantitative Features. Each section has a dropdown menu to select features. The Categorical Features section shows 'state', 'area_code', 'intl_plan', and 'vmail_plan' selected. The Categorical Array Features section shows a dropdown menu with the text 'Column names which should be treated as categorical array data'. The Quantitative Features section shows 'vmail_message', 'day_mins', 'day_charge', 'eve_mins', 'eve_charge', 'night_mins', 'night_charge', 'intl_mins', 'intl_charge', and 'custserv_calls' selected.

2. Select **Save** and **Train** on Predictive Scoring to build the predictive model.



The screenshot shows a user interface with two buttons: 'SAVE' and 'TRAIN'. The 'SAVE' button is grey and the 'TRAIN' button is blue. The buttons are positioned below a dashed line and a solid line.

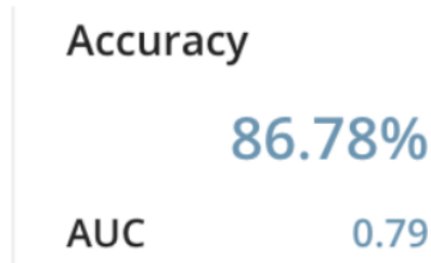
3. Select Predictive Scores to review the score distributions.



Interpreting the Accuracy of Prediction

When you train Predictive Scoring, an evaluation runs in parallel to estimate the accuracy of prediction.

The metrics of Accuracy and AUC ([Area Under the ROC Curve](#)) can help you determine how to tune your predictive scoring.



You can interpret the value of AUC as follows:

- AUC between 0.7 and 0.9 is optimal.
- If AUC is less than 0.7, you should refine your predictive model to improve its accuracy.
- If AUC is greater than 0.9, consider the possibility of [data leakage](#).

The possibility of leakage occurs when you choose attributes that are directly related to the solution to your problem. When you use a solution for prediction, it is kind of like cheating. For example, in the [predictive scoring tutorial](#), the attribute `churn` must not be incorporated into columns configured for Predictive Scoring.

Low AUC Troubleshooting

When AUC is low, less than 0.7, it is usually because the quality of input to the machine learning algorithm is poor. You can exercise one of the following options to improve AUC:

- Add more predictive features
- Rethink your problem (such as, definition of population, positive samples, and scoring target segments)
- Revisit your master segment definition (such as, customer data), use more data, and add predictive features

Adding more features, requires that you edit the configuration in Predictive Scoring:

SUGGEST PREDICTIVE FEATURES

Categorical Features

state × area_code × intl_plan × vmail_plan ×

Categorical Array Features

Column names which should be treated as categorical array data

Quantitative Features

eve_mins × night_charge × day_charge × night_mins × custserv_calls × intl_mins × day_mins × intl_charge × eve_charge × vmail_message ×

> Feature Customization

Suggest Predictive Features automatically drops some “likely to be meaningless” attributes, but misclassification can still occur. Therefore, try to add additional columns, which are likely to be informative, to the input boxes based on your understanding of the data.

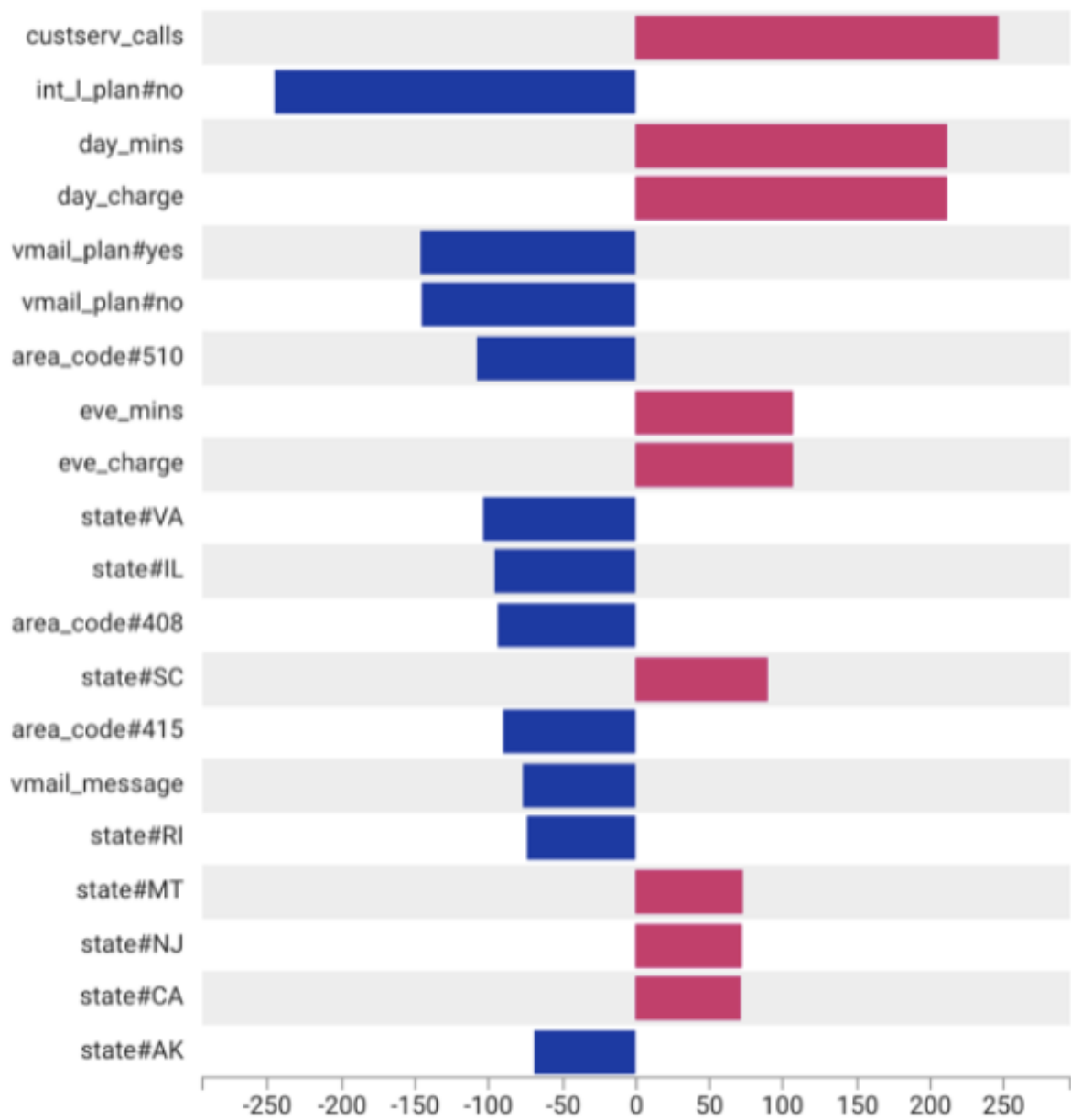
If the AUC is still less than 0.7, your master segment might not have reasonably informative attributes or the number of profiles might be too small. To resolve this you can import more data to TD and integrate it with the master segment as additional behaviors or attributes.

You can also reconsider the definition of population, positive samples, and scoring target segments. For example, if all profiles are used for building a predictive model, consider narrowing down your focus to a specific population segment.

AUC Higher than 0.9 Troubleshooting

Exceptional accuracy might occur as a result of the choice for the predictive features. Check the feature importance on the dashboard, and determine if the absolute values of importance for top features are unusually large:

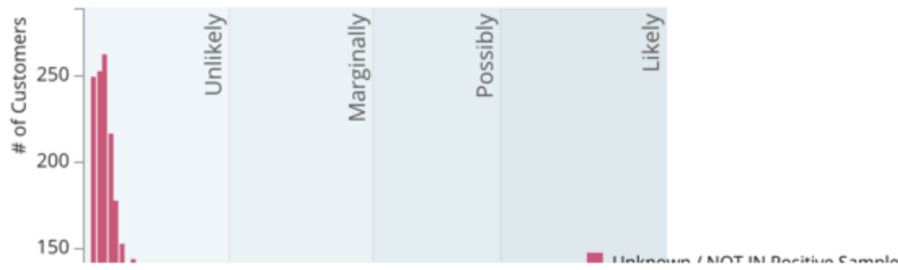
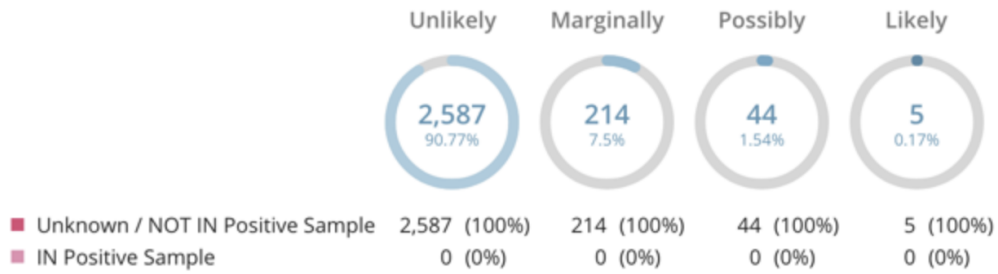
Features (Top 20)



To tune predictive scoring when AUC is greater than 0.9

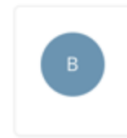
1. Drop a predictive feature corresponding to the most important feature on the Predictive Scoring.
2. Select **Save** and **Train** on Predictive Scoring to rebuild the predictive model.
3. Select Predictive Scores to review the score distributions.
4. Repeat these steps until the AUC is between 0.7 and 0.9.

Score Distribution



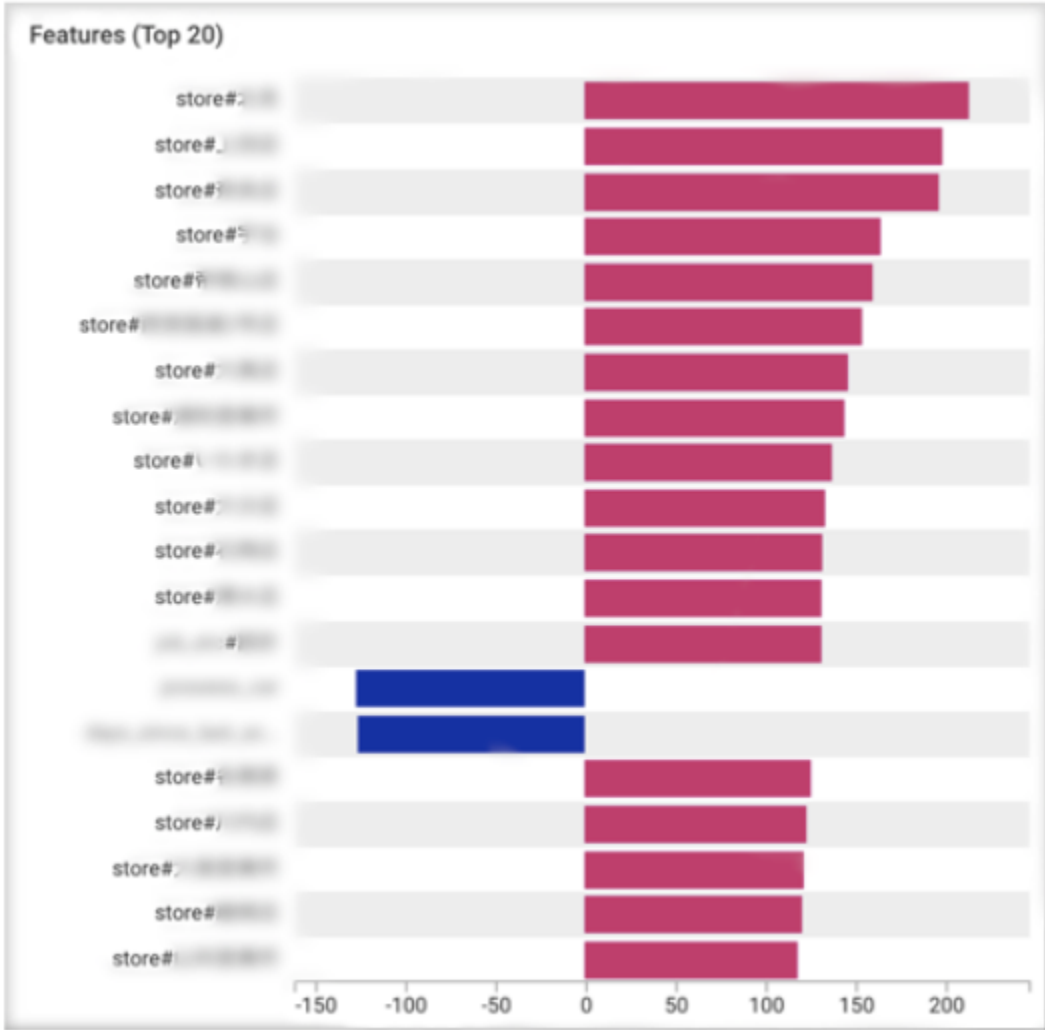
Profiles

Likely



Possibility of Overfitting

The reason for very low or exceptionally high accuracy is sometimes overfitting, sometimes referred to as overtraining. Overfitting occurs when the prediction model memorizes the data points in its entirety rather than by learning patterns. For example, too many **store#**.



The predictive score for a profile that has specific value in an attribute `store` is likely to become large, regardless of values in the other attributes. Consequently, predictive scores become meaningless across a master segment, and accuracy can become strangely biased.

If you observe overfitting when you view the dashboard, we recommend that you drop the overfit column on Predictive Scoring and rebuild your predictive model.