Lab 3b: Classification\_Model

Testing & Validation

Objectives

In this lab you will evaluate two different models previously created using model accuracy and financial performance evaluation techniques.

* Approximate Completion time: **25 minutes**

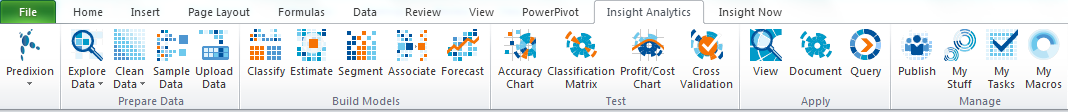
Prerequisites

* Have materials folder available to navigate to the appropriate datasets.

# Accuracy and Chart Explained

### Accuracy Chart

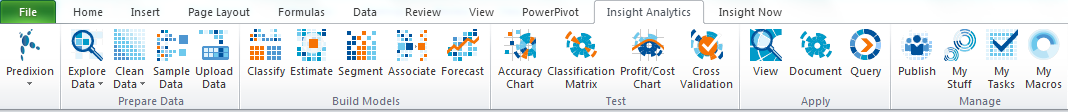
Next we will compare the results of the two models using the Accuracy Chart option located on the Insight Analytics Ribbon under the Test category. The Accuracy chart helps you evaluate the performance of predictive models against the model test data or other data you choose.



### Profit/Cost Chart

The Profit Chart wizard helps you build a chart that does two things: it measures the accuracy of your model, and it uses costs that you provide to generate a profit curve that visually represents the potential worth of a predictive model.

Whereas the accuracy chart assesses only the probability that predictions are right or wrong, the profit chart incorporates real-world knowledge about the consequences of taking action on a prediction. For example, it can visually demonstrate the projected increase in profit when you use a model, or the costs of making a prediction that is wrong. Incorporating costs can dramatically affect the way that you make decisions, such as which customers to approach. In this lab we will be using the thin client version of the Cost Chart to demonstrate the financial analysis of the models we built in the prior labs.



# Hands On Exercise – Model Testing & Validation

|  |  |
| --- | --- |
| **Tasks** | **Action** |
| 1. Create an Accuracy Chart. | 1. Select **ReadmissionsSanitizedTraining.xlsx** and **Open** to access the file. Select the RawDataSet tab 2. Select the **Accuracy Chart** button from the Insight Analytics ribbon. Select **Next**.   C:\Users\KSOUTH~1\AppData\Local\Temp\SNAGHTML964ba79.PNG   1. From the Readmission Dataset or Model page, make sure you select the **Readmission Dataset** as pictured below and select **Next** to compare the models (and not a specific Model).      1. Insure the Mining Model to predict is set to **Readmitted** and that the **Value to predict** is set to **1** then select **Next**.      1. Leave the default selection of **Model Test Data** and then select **Finish**.   C:\Users\KSOUTH~1\AppData\Local\Temp\SNAGHTML965449f.PNG |
| 1. Open and interpret the Accuracy Chart. | 1. After the job has completed the report will be added as a new sheet in your workbook. You should see a tab named **Accuracy Chart** (if not, please select **Results** to open the tab). 2. Review the findings in the chart to determine which model is the best performer.     Lift  Lift is a measure of the effectiveness of a predictive model calculated as the ratio between the results obtained with and without the predictive model.   * Cumulative gains and lift charts are visual aids for measuring model performance * Both charts consist of a lift curve and a baseline * The greater the area between the lift curve and the baseline, the better the model   Cumulative Gains Chart   * The y-axis shows the percentage of positive responses. This is a percentage of the total possible positive responses (as the overall response rate shows). * The x-axis shows the percentage of customers contacted, which is a fraction of the total customers. * Baseline (overall response rate): If we contact X% of customers then we will receive X% of the total positive responses. * Lift Curve: Using the predictions of the response model, calculate the percentage of positive responses for the percent of customers contacted and map these points to create the lift curve.   Lift Chart   * Shows the actual lift. * To plot the chart: Calculate the points on the lift curve by determining the ratio between the result predicted by our model and the result using no model. * Example: For contacting % of customers, using no model we should get % of responders and using the given model we should get 30% of responders. The y-value of the lift curve at % is 30 / = 3.   Analyzing the Charts  Cumulative gains and lift charts are a graphical representation of the advantage of using a predictive model to choose which customers to contact. The lift chart shows how much more likely we are to receive respondents than if we contact a random sample of customers.  Evaluating a Predictive Model  We can assess the value of a predictive model by using the model to score a set of customers and then contacting them in this order. The actual response rates are recorded for each cutoff point, such as the first % contacted, the first 20% contacted, etc. We create cumulative gains and lift charts using the actual response rates to see how much the predictive model would have helped in this situation. The information can be used to determine whether we should use this model or one similar to it in the future.  **Which Model is the best? Why?** |

|  |  |
| --- | --- |
| **Tasks** | **Action** |
| 1. Perform Cost Analysis | 1. Select the **View** button from the Insight Analytics ribbon. This will bring up a dialog box that lets you choose a model to perform cost analysis in a thin client.      1. Select **Classified Readmitted – Logistic Regression model (Or something like: Classified Readmitted\_LogReg)** 2. Click the OK button. 3. This will bring up a thin client web view as shown below. You may need to resize the thin client viewer. |
| 1. Open the Accuracy view | 1. On the far left side of the thin client is a frame that lets you choose different views. Select the **Accuracy** **Chart** view. This should bring up a view as shown below. 3. Set the “Accuracy of predicting Readmitted = 1” as shown below. This is done by pulling down on the **0** in the on the label “Accuracy of Readmitted = 0 at the top of the screen. 5. Select the **Cost** Chart from the list of charts in the thin client. This thin client view should change as shown below. 7. Set the cost parameters. In the upper right area of the Cost chart there is a gear icon. Click on this icon and the list of parameters will appear that can be edited. Enter the following values: 8. **Population**: 11,009 9. **Fixed Cost**: $ 0 10. **Individual Cost**: $1,000 11. **Undetected Cost**: $20,000 12. **Efficacy:** .3 14. Select the list icon next to the gear icon on the thin client view and deselect all check boxes and then check the following checkboxes: 15. **Minimum Cost** 16. **Minimum Threshold** 17. **ROI** 18. **Savings** 19. Click in the graph to close the list checkbox. 20. Here is what the chart is telling us. If we have a patient population of 11,009 members and assuming that it costs $1,000 to intervene on patients that we predict will be readmitted versus doing nothing at a cost of $20,000 (on average) and assuming an efficacy value of 30%, we can expect the cost savings to be $4,095,048 and the ROI to be 124%. This savings would be achieved setting the minimum threshold at 26%. |
|  | In the upper right corner, select the check boxes for the other models in this dataset like **Classified Readmitted\_NN, Classified Readmitted\_NB, Classified Readmitted\_DT** (Neural Net, Naïve Bayes, and Decision Trees)  Take a screen shot of all the models and submit to the Catalyst site. |

## End of Lab

Please proceed to the next lab.