BigML for Alexa Manual

The BigML Team

Version 1.0

MACHINE LEARNING MADE BEAUTIFULLY SIMPLE

About this Document

This document gives an introduction to the BigML for Alexa skill which brings voice control to BigML's predictive capabilities.

Contents

What is the BigML for Alexa Skill?

1.1 Overview

The BigML for Alexa skill combines the predictive power of BigML with the voice processing capabilities of the Alexa Voice Service. Using an Alexa enabled device like an Amazon Echo or Dot, this integration makes it possible to use spoken questions and answers to generate predictions using your own models trained in BigML.

For example, if you have data regarding wine sales with features like the sale month and grape variety, you could build a model which predicts the sales for a given month, variety, etc. With this model loaded into the BigML for Alexa skill, you could generate a sales prediction by answering questions vocally.

Figure 1.1: Overview of the BigML for Alexa Skill

The significance of the BigML for Alexa skill is subtle. While any predictive platform could probably be integrated with a voice control system, BigML's Machine Learning algorithms are especially well suited to the task because of two reasons:

- BigML's algorithms generate models which can make predictions using an optimal path. This means that if you have a dataset with, say, thousands of features, it is possible that only a handful of answers are all that is required to create a prediction. This capability is essential in a voice control interface since it would be tedious to answer thousands of questions vocally.

- BigML's algorithms can handle missing data. This means that you do not need to know the answer to every question. The ability to answer "I don't know" to a voice prompt is very powerful, especially if the prediction could not otherwise proceed until you provided every detail, even if not significant.

For now, the BigML for Alexa skill is focused on making predictions, but a future version will likely support the ability to perform other Machine Learning tasks by voice control as well, especially with the ability to run workflows using WhizzML. If you have any feedback or suggestions for the BigML for Alexa skill, be sure to send them to us at mailto:info@bigml.com.

1.2   Quick Start

To quickly get a feel for the capabilities of the BigML for Alexa skill simply enable the skill with:

"Alexa, enable the Big M. L. skill"

And then ask to run a demo:

"Alexa, ask Big M. L. to give me a demo"

This will load a model which asks questions about a patient's diagnostic measurements like a 4-hour plasma glucose and BMI and uses your answers to make a prediction about the likelihood of that individual having diabetes.

Note: This is only a demo and is not medical advice! The demo model was built from the small UCI dataset found here https://archive.ics.uci.edu/ml/datasets/Diabetes and no effort was made to optimize or evaluate the model for accuracy.

1.3   Example Demo Session

Here is an example session, using the demo mode.

"Alexa, ask Big M. L. to give me a demo"

"Welcome to the BigML Demo."

"I have loaded a model which makes predictions about diabetes."

"Would you like to start a prediction?"

"Yes"

"What is the patients' 4-hour plasma glucose?"

"One hundred eighty"

"What is the patient's body mass index?"

"Thirty five"

"What is the patient's diabetes pedigree?"

"One"

"The prediction for diabetes is true, with a confidence of 91 percent"

Getting Started

Getting the BigML for Alexa skill setup to work with your own models is a four step process:

1. Enable the BigML for Alexa skill

2. Link your BigML account

3. Load your own predictive models

4. Make predictions

2.1 Enabling the BigML for Alexa Skill

The easiest way to enable the BigML for Alexa skill is to simply say:

> *"Alexa, enable the Big M. L. skill"*

Or you can enable the skill by first searching for it in the "Skills" section of the Alexa App and enable it directly:
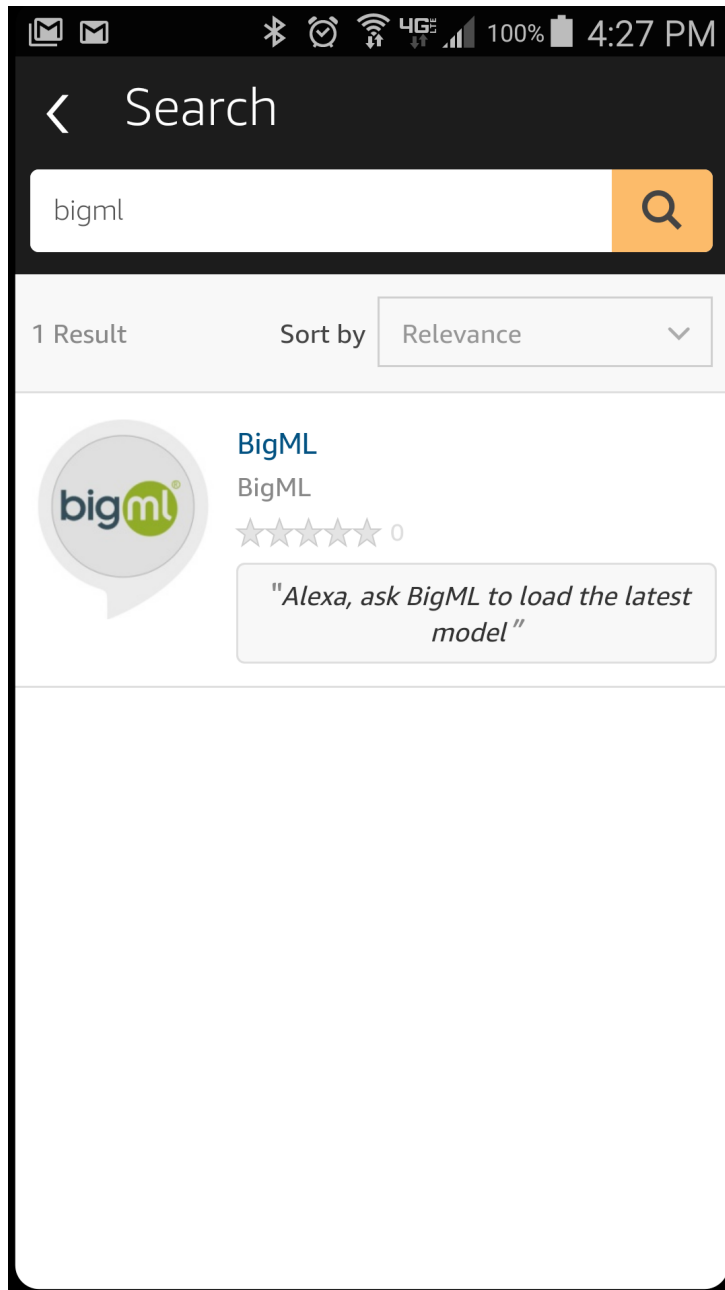
Figure 2.1: Searching for the BigML Skill in the Alexa App
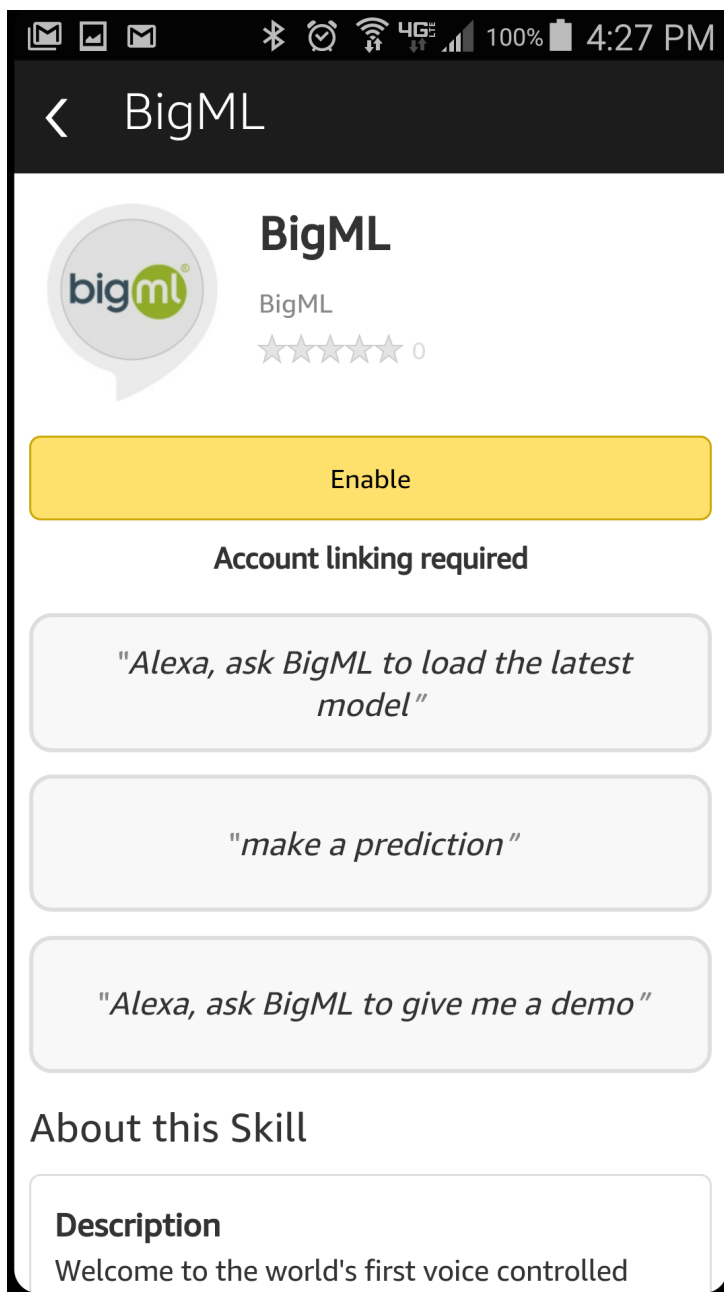
Figure 2.2: Selecting the BigML Skill in the Alexa App

2.2   Linking to Your BigML Account

Once the BigML for Alexa skill is enabled, you need to link the skill to your BigML account to be able to load your own models. If you do not already have a BigML account, you should create one now by visiting https://bigml.com/accounts/register/.

If you just enabled the BigML for Alexa skill, it will automatically start the linking process. If not, you can open the BigML for Alexa skill in the "Skills" section of the Alexa app and then simply press "Link Account". Or you can say:

"Alexa, open Big M. L."

You will hear a welcome message after which the skill will send a card to your Alexa app with the url to complete account linking.

Figure 2.3: Linking the BigML Skill in the Alexa App

Once the account linking screen is open, simply enter your BigML username and password. If the process is successful, you will see the following screen:
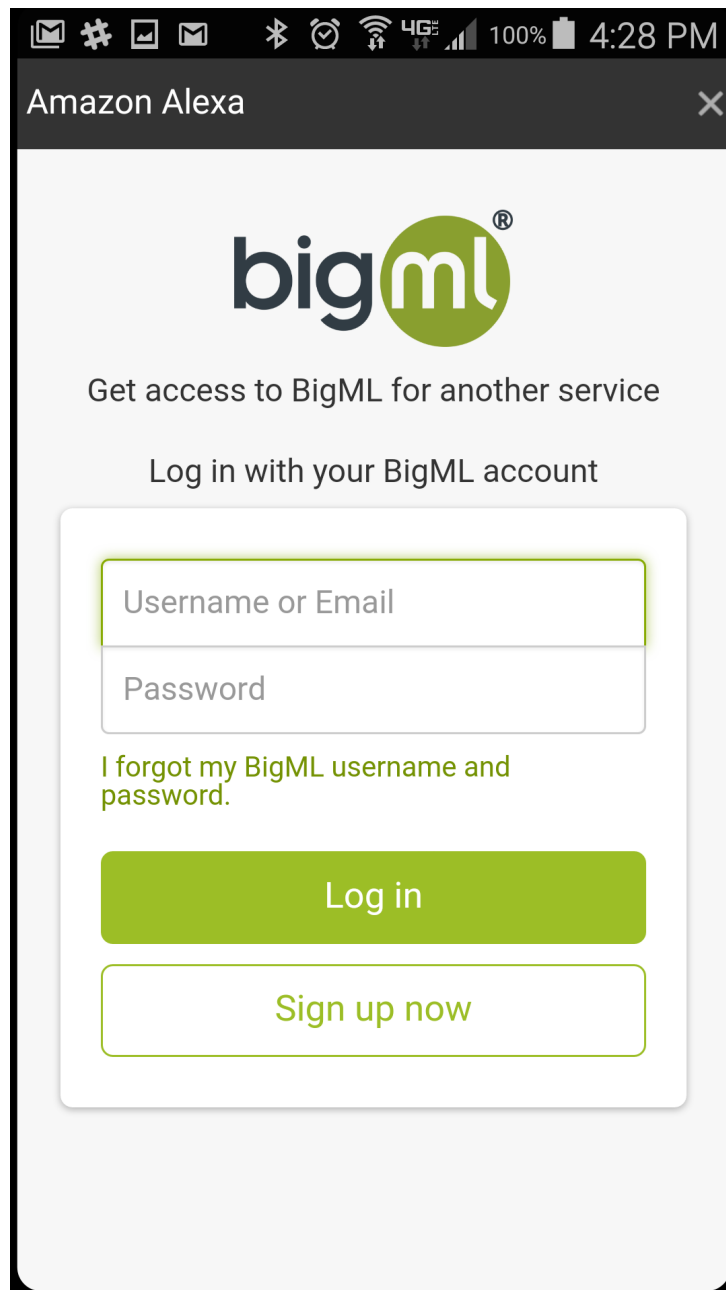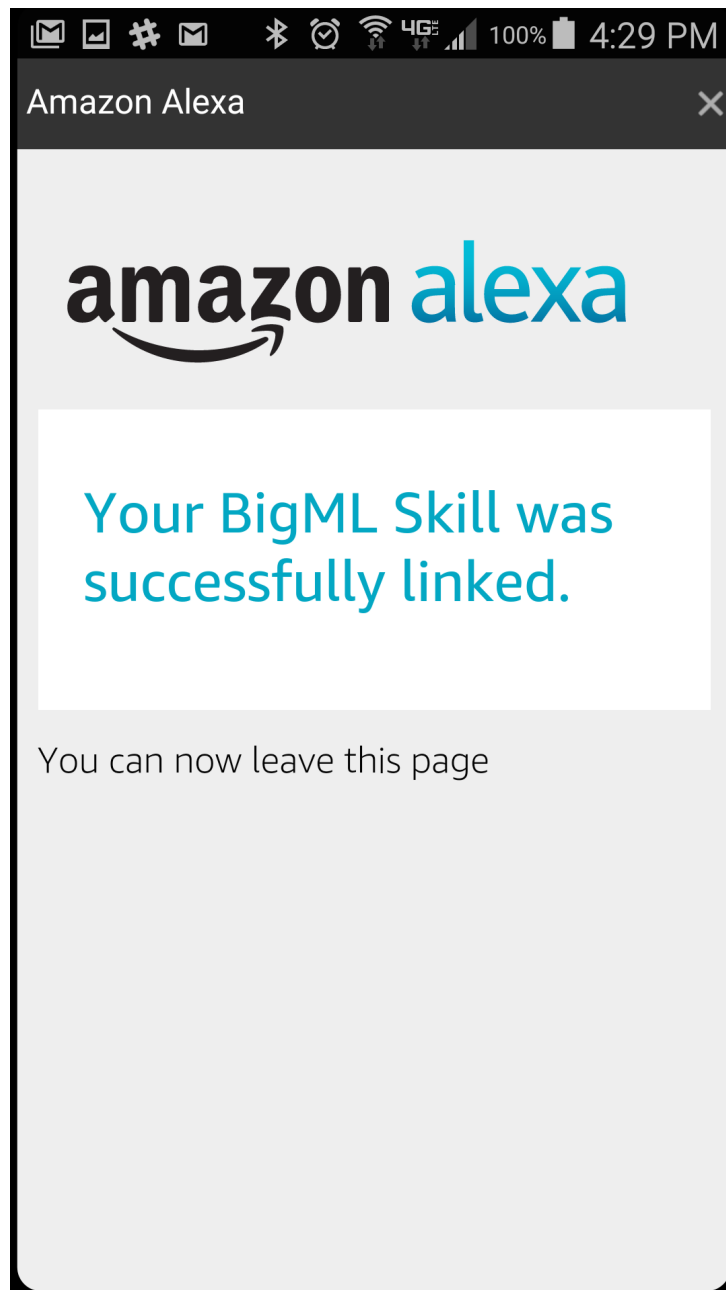
Figure 2.4: Successful Link of the BigML Skill in the Alexa App

2.3   Loading Models

Once the BigML skill is enabled and linked to your BigML account, you can make voice predictions using the models in your BigML account. Before you can make the first prediction for a given model, you need to "load" the model into the BigML for Alexa skill service. The voice command to load the latest model in your account is:

> *"Alexa, ask Big M. L. to load the latest model"*

This will instantiate the most recently created model from your BigML account into the BigML for Alexa service and prepare it for making predictions. The BigML for Alexa skill will then ask you if you would like to start a prediction.

You can also load ensembles, clusters, and anomaly detectors although these are not supported for predictions yet.

2.4   Making a Prediction

After the above steps have been completed, you are ready to make a prediction. The command to start a prediction is:

"Alexa, ask Big M. L. to make a prediction"

The BigML for Alexa skill will use the structure of your model to figure out the first question to ask. Subsequent questions will depend on the answers you give as the skill finds the optimal path to make the prediction.

2.5   Example Session

Here is an example session, using the wines sales model.

"Alexa, ask Big M. L. to load the latest model"

"The latest wine sales model is loading."

"Would you like to start a prediction?"

"Yes"

"What is the price?"

"Sixteen"

"What is the grape?"

"Pinot Noir"

"Is the wine from Oregon?"

"Yes"

"The prediction for total sales is one hundred fifty."

Tips and Tricks

Now that you have the basics of the BigML for Alexa skill, there are several tips and tricks that are useful to know.

3.1 Project Filters

While you can currently only load the "latest" model, you can filter the models using projects in your BigML account. For example, you could have separate projects for "Sales" and "Resumes" and then to load the latest model from the "Sales" project you would first set the project:

"Alexa, ask Big M. L. to set the project"

"What is the project name?"

"Sales"

"Selected project Sales"

"Would you like to load the latest model?"

"Yes"

Once a project is set, the BigML for Alexa service will remember it and load all resources from only that project. This means you can create a new model in the "Sales" project and then simply ask to "load the latest model".

If you would like to clear the project filter, you can do so by setting it to "all":

"Alexa, ask Big M. L. to set the project"

"What is the project name?"

"All"

"The project filter has been removed."

"Would you like to load the latest model?"

"No"

3.2 Controlling Field Questions

When making a prediction, the BigML for Alexa service has to figure out how to phrase the question for each input required. This is especially important if your dataset has field names that are not speakable.

For example, you might have a dataset of loan information with a field named "int_rate" but you would like the BigML for Alexa skill to ask "What is the interest rate" instead of "What is the int underscore rate".

It is possible to control the spoken definition for each field name using the **label** and **description** meta-data in the dataset in BigML.

The skill uses the following rules to build the question for a given field in the dataset. The rules are listed in their priority order, that is, if a field has a description, then the label will not be used to build the question, etc:

description If the field name has a description, it's value will be used explicitly for the field question. This give you complete control over how the field will be asked for.

built-in If the field name matches a set of built-in patterns, then the internal map will be used for the field question. See Section 3.10 for a list of current built-in prompts.

label If the field has a label, then the question will be "What is the field_label?"

name If none of the above rules matches, then the question will be "What is the field_name?"

The exception to the above rules is the "items" data-type. Since this input field requires a list of values for a single input, the voice prompt will always be: "I will ask for field_name values one at a time. Say stop to end. What is the field_name"

After each valid item, the skill will reprompt with "What is the field_name" until you say "stop".
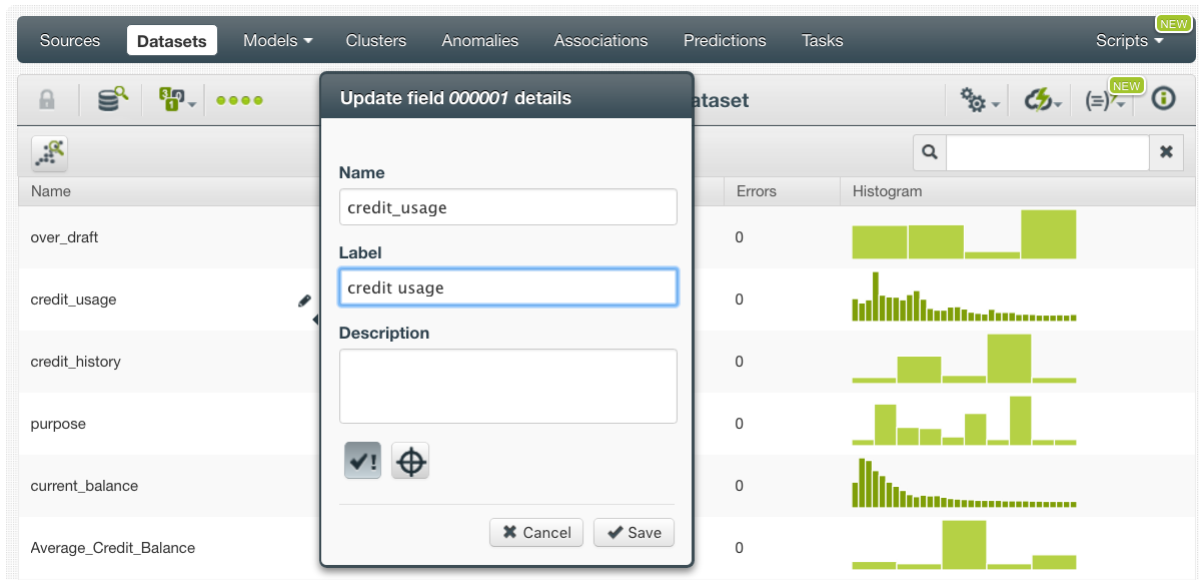

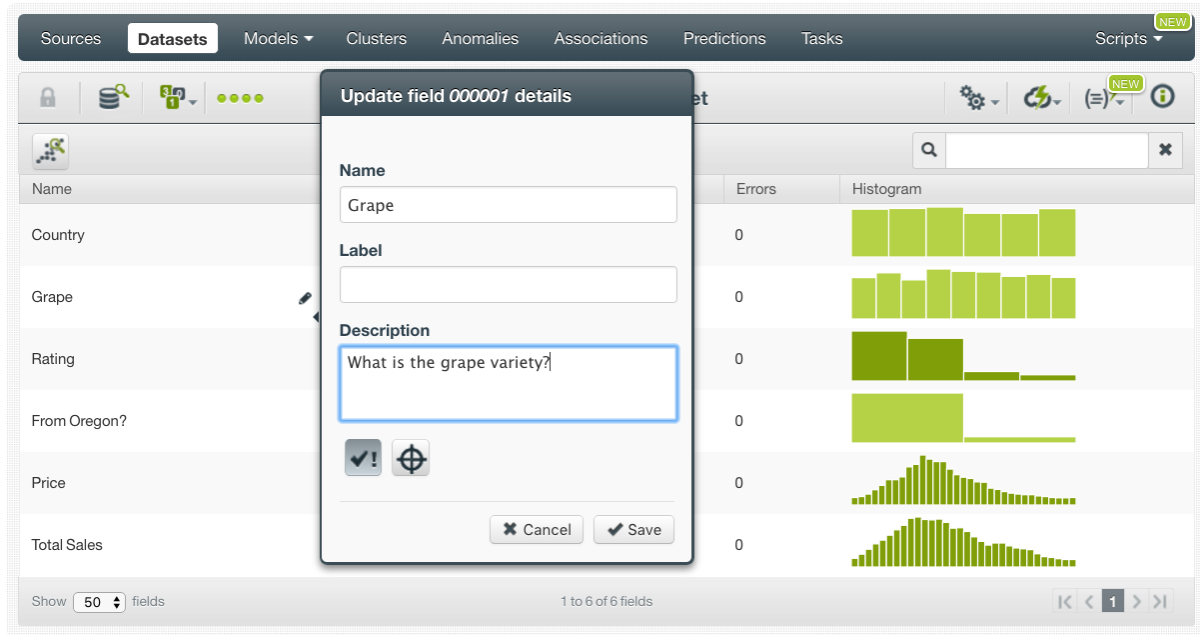Figure 3.1: Correcting a field name using the field label

Figure 3.2: Creating a custom field question using the field description

3.3   Controlling Field Statements

When a prediction is complete, the BigML for Alexa service has to figure out how to phrase the prediction statement for the final output. As with Section 3.2 this is important if your model objective has a field name that is not speakable.

It is possible to control the spoken prediction statement using the **label** and **description** meta-data in the dataset in BigML.

The skill uses the following rules to build the prediction statement for the objective. The rules are listed in their priority order, that is, if a field has a description, then the label will not be used to build the statement, etc:

description If the field name has a description, it's value will be used explicitly for the field statement. This give you complete control over how the prediction will be stated.

built-in If the field name matches a set of built-ins patterns, then the internal map will be used for the field statement. See Section 3.10 for a list of current built-in prompts.

label If the field has a label, then the statement will be "The prediction for field_label is:"

name If none of the above rules matches, then the statement will be "The prediction for field_name is:"
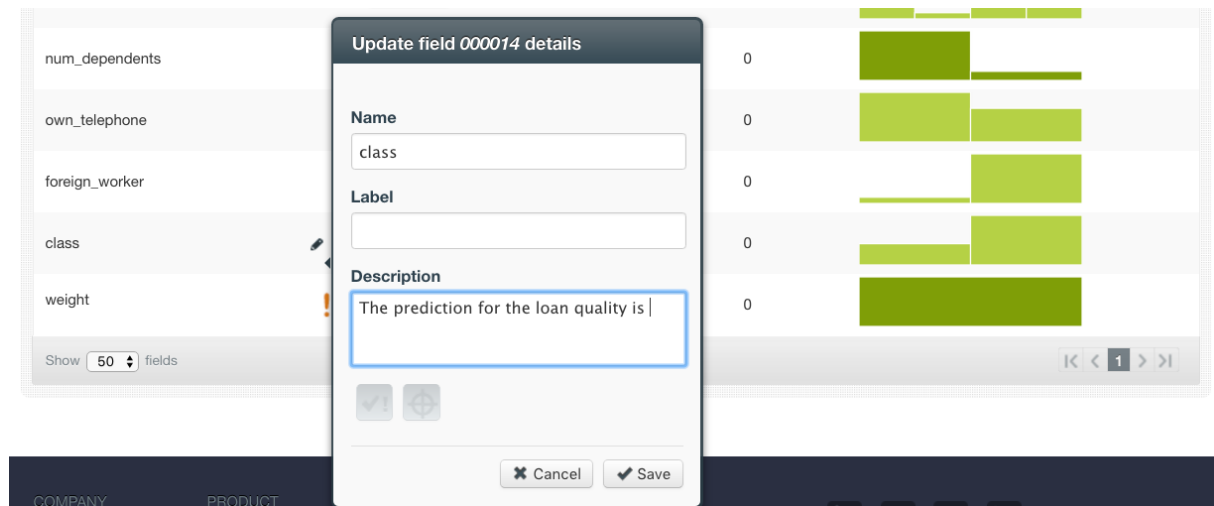
Figure 3.3: Creating a custom field statement using the field description

3.4 Reprompts

The BigML for Alexa skill makes use of reprompts to provide additional information when a response is needed. The reprompts are issued after 10 seconds of waiting for an answer.

At prediction time, these reprompts will contain examples of valid inputs for the field that is being prompted. You will also hear these valid input examples when the BigML for Alexa skill hears an invalid response.

Reprompt Example:

"Alexa, ask Big M. L. to load the latest model"

"The latest wine sales model is loading."

"Would you like to start a prediction?"

"... no answer ..."

"If you would like to start a prediction, just say yes"

"Yes"

"What is the price?"

"... no answer ..."

"The price should be a number: expected values are between"

"6.95 and 34.99. You can also say stop to end input and make"

"a prediction. Or you can say cancel to reset the questions."

"Elephant"

"The price should be a number: expected values are between"

"6.95 and 34.99. You can also say stop to end input and make"

"a prediction. Or you can say cancel to reset the questions."

3.5   Answering with Unknown

When making a prediction, you do not have to provide an answer to every question. If you do not know an answer, you can simple say "I do not know". The BigML for Alexa skill will ask if you want to skip this question. If you answer "yes", the skill will mark this question as unknown and then continue asking for other inputs to make the best prediction possible. If you answer "no" then the skill will ask the input question again.

Unknown Example:

"Alexa, ask Big M. L. to make a prediction"

"What is the price?"

"I do not know"

"Would you like to skip this prediction question?"

"Yes"

"What is the grape?"

3.6   Stopping

For most parts of the BigML for Alexa skill, saying "stop" will stop the current action and close the session. However, there are two situations where the stop intent behaves differently:

1. When answering questions to make a prediction, saying "stop" will mark all remaining questions as unknown and generate a prediction immediately.

2. When providing individual input values for an input field with the "items" data type, saying "stop" finalizes the list of items and continues with the next field question.

Prediction Stop Example:

"Alexa, ask Big M. L. to make a prediction"

"What is the price?"

"Stop"

"The prediction for total sales is 76."

Items Stop Example:

"Alexa, ask Big M. L. to make a prediction"

"I will ask for cart values one at a time."

"Say stop to end. What is the cart?"

"Ice Cream"

"Next cart item?"

"Coffee"

"Next cart item?"

"Stop"

"The prediction for the month is seven"

3.7   Canceling

For all parts of the BigML for Alexa skill, saying "cancel" will stop the current action and close the session. In addition, if you say "cancel" when answering questions to make a prediction, then the BigML skill will also abandon the prediction before closing the sessions. Note that this behavior is different from saying "stop" in that the prediction is not made, not stored in your BigML account.

Cancel Example:

"Alexa, ask Big M. L. to make a prediction"

"What is the price?"

"Cancel"

"Would you like to cancel this prediction?"

"Yes"

"Canceled"

3.8   Unlinking the Skill

If you would like to unlink your BigML account from your Alexa device, simply disable the skill. This automatically removes the authentication token and unlinks your account. You can then re-enable the skill and authenticate with a different BigML account.

3.9   Data Type Detection

When making a prediction, the BigML for Alexa skill listens to the response and from what is heard has to figure out the data type, that is whether what you said was an integer, real number, random text, categorical, etc. It then matches the data type it heard against the data type of the input field it is asking for, and if they match the input is accepted, otherwise it will reprompt for the field.

In general, the BigML for Alexa skill is very good at distinguishing spoken integers, like "one hundred" and in fact is also good at getting colloquial numbers correct like "one eighty" meaning 180.

When providing a real number, there are a few ways to speak the number to ensure that it is heard correctly. For example. to provide the input "3.14", these utterances should all work:

"three point one four"

"three dot one four"

"three and fourteen"

For categorical values, that is inputs that have a fixed list of valid inputs, the BigML for Alexa skill will do it's best to interpret what it hears, and then will compare to the list of valid inputs. The comparison is done insensitive to case and also uses metaphones to compare using word sounds as well.

For example, if an input value is categorical and the possible values in your BigML dataset are, say "Shoe" and "Socks" and the BigML for Alexa skill hears "shoo" then it will pass the value "Shoe" to your model to make the prediction.

The skill will also accept yes/no responses for categorical inputs that include true/false values.

The BigML for Alexa skill even accepts text input, although the timeout for devices like the Amazon Echo make inputting lengthy text impractical.

It is important to keep this data type detection in mind when structuring your dataset. For example, if you have a categorical input with values like:

["10 Percent", "20 Percent", ...]

Then when asking for this input, the BigML for Alexa skill will have trouble matching the spoken values to the list because they partially match both numbers and text. In this case it would be better to remove the "Percent" text from each label. This will allow the skill to detect a spoken integer and then match the number correctly.

A similar problem would exist if your field values were the spoken numbers:

["Ten", "Twenty", ...]

In this situation, the BigML for Alexa skill will interpret the spoken numbers as integers and then will not be able to match "10" to the label "Ten". In this case, it would be better to replace the spoken values with the actual numbers, or perhaps to even convert this field to a numeric type.

It is also important to watch for symbols and punctuation, for example having a categorical value like "set_a". When heard, the symbols will be converted into the spoken representation, in this case "set underscore a" and will not match the label value.

We are constantly improving the capabilities of the BigML for Alexa skill with regard to correct data type recognition. If you run into a situation that you can not work around or that does not seem to behave correctly, let us know at `mailto:info@bigml.com`

! As a rule of thumb, keep your input types as simple as possible, prefer integers and real numbers over categorical or text and remove symbols and punctuation.

3.10   Built-In Prompts

The following tables list all of the current built-in prompts. If the field name in your dataset matches one of the entries below and you do not have custom settings defined, then the BigML for Alexa skill will use the built-in question and label to construct the field question and statement.

These built-in prompts have been added to make demoing with a few UCI datasets easier, but can also be useful for your own dataset. If you have a suggestion for a common field name that we should add to the built-ins, please let us know at `mailto:info@bigml.com`.

Field Name	Label	Question
pregnancies	pregnancies	what is the number of times the patient has been pregnant
plasma glucose	plasma glucose	what is the patients four hour plasma glucose
blood pressure	blood pressure	what is the patients blood pressure
triceps skin thickness	triceps skin thickness	what is the patients triceps skin fold thickness
insulin	insulin	what is the patients two hour serum insulin level
bmi	body mass index	what is the patients body mass index
diabetes pedigree	diabetes pedigree	what is the patients diabetes pedigree
age	age	what is the patients age
diabetes	diabetes	what is the patients diabetes assessment

Table 3.1: Built-In Prompts for UCI Diabetes

Field Name	Label	Question
zip	zip code	what is the zip code
beds	bedrooms	what is the number of bedrooms
baths	bathrooms	what is the number of bathrooms
sqft	square footage	what is the square footage
parking spots	parking spots	how many parking spots

Table 3.2: Built-In Prompts for RedFin

Field Name	Label	Question
loan_amount	loan amount	what is the loan amount
int_rate	interest rate	what is the interest rate
sub_grade	sub grade	what is the sub grade
emp_title	job title	what is the job title
home_ownership	home ownership	does the applicant own a home
annual_inc	annual income	what is the annual income
is_inc_v	income verification	is the income verified
delinq_2yrs	delinquencies in two years	What is the Number of delinquencies in the last two years
inq_last_6mths	six month inquiries	what is the Number of credit inquiries in the last six months
open_acc	open accounts	how many open accounts does the applicant have
pub_rec	public records	how many of public records
revol_bal	revolving balance	what is the revolving credit balance
revol_util	revolving utilization	how much revolving credit is utilized
total_acc	total account	what is the total account balance
total_rec_late_fee	total late fees	what is the total of late fees received
collections_12_mth	collections in last year	how many collections has the applicant had in the last year

Table 3.3: Built-In Prompts for Lending Club

Field Name	Label	Question
intl plan	international plan	does the customer have an international plan
vmail plan	voicemail plan	does the customer have a voice mail plan
nmb vmail mssg	number of voicemails	how many voice mail messages were received
ttl day min	total daytime minutes	What is the total number of daytime minute used
ttl day calls	total daytime calls	How many daytime calls were made
ttl day charge	total daytime charge	What is the total daytime call charge
ttl eve min	total evening minutes	What is the total number of evening minute used
ttl eve calls	total evening calls	How many evening calls were made
ttl eve charge	total evening charges	What is the total evening call charge
ttl nght min	total night minutes	What is the total number of night time minute used
ttl nght calls	total night calls	How many night time calls were made
ttl nght charge	total night charges	What is the total night time call charge
ttl intl min	total international minutes	What is the total number of international minute used
ttl intl calls	total international calls	How many international calls were made
ttl intl charge	total international charges	What is the total international call charge
cust serv calls	customer service calls	How many calls for customer service were made

Table 3.4: Built-In Prompts for Telecom Churn

List of Figures

List of Tables