BigML Zapier App

The BigML Team

Version 1.0



MACHINE LEARNING MADE BEAUTIFULLY SIMPLE

Copyright © 2018, BigML, Inc.

Copyright© 2018, BigML, Inc., All rights reserved.

info@bigml.com

BigML and the BigML logo are trademarks or registered trademarks of BigML, Inc. in the United States of America, the European Union, and other countries.

This work by BigML, Inc. is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. Based on work at http://bigml.com.

Last updated June 14, 2018

About this Document

This document describes the BigML Zapier app, which aims to make it easy to build automated Machine Learning workflows integrating external services and applications with BigML.

Contents

1	Introduction	1
2	Getting Started2.1Setting Up the Trigger: Google Sheets App2.2Setting Up the Action: BigML Zapier App2.3Sending an E-Mail with the Prediction: Gmail App2.4Turning Your Zap On	7 13
3	BigML Zapier App Actions3.1Set Up Common Arguments for All Actions3.2Create Prediction3.3Create Centroid3.4Create Anomaly Score3.5Create Topic Distibution3.6Find a Resource	26 28 28
Lis	st of Figures	32
Gl	ossary	34
Re	eferences	36



Introduction

Zapier¹ is a web service that allows you to **create workflows** using different web services such as Google Drive, Salesforce, Gmail, Facebook, and many others among its 1,000+ applications².

The BigML Zapier app³ allows you to **make predictions** using a model, ensemble, logistic regression, deepnet, cluster, anomaly, or topic model as part of a larger Zapier workflow. We describe two scenarios below as illustrative examples:

- An IoT device for healthcare monitors some patient's health measurements such as their insulin, glucose, and blood pressure levels. It sends periodically, e.g., every hour, the measured data to some remote service, e.g., Google Sheet or a database. Whenever new data comes in, a Zapier trigger associated to that service reads it and passes it to the BigML Zapier app, which will predict the likelihood of diabetes. The prediction outcome is then used to trigger sending an email to warn the doctor or patient in case the confidence of a diabetes diagnosis is high. This example is described step-by-step in Chapter 2.
- An e-commerce service stores all processed orders in Salesforce, along with the data about the buyer, the payment, and any other significant information to describe the transaction, such as whether the delivery was disputed, the product was returned, a refund required, etc. For each new order coming in, you could trigger a prediction using a BigML model to evaluate the likelihood of that transaction to fail for any reasons. If the prediction confidence is higher than a given threshold, you can use another Zapier action to flag the transaction in Salesforce as requiring ad-hoc tracking by a human controller.

These two examples and many other BigML integrations⁴ can be developed by creating a Zapier workflow and using the BigML Zapier app without requiring **any coding skills**.

Note: this document does not attempt to be a guide to the Zapier platform. It aims to provide some easy step-by-step examples that do not assume you are a Zapier advance user, but you need to understand the basics of how Zapier works⁵, e.g. what a Zapier workflow (Zap⁶) is and how to build one by connecting triggers and actions from multiple apps together.

³https://zapier.com/apps/bigml/integrations

¹https://zapier.com

²https://zapier.com/apps

⁴https://zapier.com/apps/bigml/integrations

⁵https://zapier.com/help/how-zapier-works/

⁶https://zapier.com/help/zaps/



Getting Started

In this chapter, you can find an easy example to use the BigML Zapier app. Imagine a patient with a risk of diabetes that is being monitored by an IoT device each hour. The device measures some health variables such as the insulin, glucose level, and blood pressure and it sends the data to a Google Spreadsheet. When the new data is generated, we want the BigML Zapier app to predict the likelihood of diabetes and then automatically send an email with the prediction.

This Zapier workflow (or Zap) has three simple steps:

- 1. Poll Google Sheets for new input data from the patient. See Section 2.1.
- 2. Use the new data to make a prediction using an existing BigML model. See Section 2.2.
- 3. Send an e-mail with the predictions. See Section 2.3.

2.1 Setting Up the Trigger: Google Sheets App

A trigger in Zapier monitors the changes in a data source. When Zapier detects new data, an action is triggered to process it. A trigger is usually the first step you need to define for any Zapier workflow.

In this section, we are going to instruct our Zap to detect when new data has been generated in our Google Spreadsheet so this data can be used to trigger a prediction in the next step (see Section 2.2).

1. Click the Make Zap button at the top of the Zapier page.

apier		Q Home	Apps Integr	ations Tips & Advice	e Invite Tea	am Make a Za	ap! TR 斗
plore Zaps	Task History Con	nected Accounts					
		Pick Ap	ops to Explo	ore Workflow	Ideas		
Q Search 100)0+ apps			1 app selected	Sort By: Popul	ar	\sim
Q Search 100	00+ apps	3	۲	(1 app selected)	Sort By: Popul	ar	PREMIUM
	00+ apps	Slack	MailChimp			Twitter	PREMIUM
			MailChimp	31		Y	PREMIUM

Figure 2.1: Create a new Zap

2. You will be redirected to the Zap creation view where the first action required is to select the Google Sheets app, which will be used as a trigger for your Zap.

$\underline{\text{Dashboard}} \ \rightarrow \ \textbf{Choose App or} \ \underline{\text{View Task History}}$	举 YOUR	ZAP IS OFF
✓ Name your zap Add a note	Choose a Trigger App	
C 1. TRIGGER	Search hundreds of apps	+
Choose App Choose Trigger	Gmail S Slack	_
Pull In Samples	MailChimp	
Set up this step	Google Calendar	
+ Add a Step	Facebook Lead Ads	
	Tunnform Narrow your search to reveal 1116 hidden results.	

Figure 2.2: Choose the Google Sheets app

3. Choose a trigger option among those provided by the Google Sheets app. Select the **New Spreadsheet Row** option, then click Save + Continue.

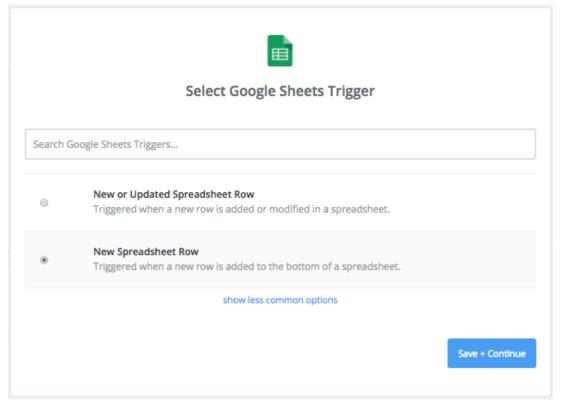


Figure 2.3: Choose the Google Sheets app as the trigger

4. Select the Google Docs account where the data is stored.

$\underline{\text{Dashboard}} \rightarrow \text{Choose Account or } \underline{\text{View Task History}}$	*	YOUR ZAP IS	OFF
Name your zap Add a note			
New or Updated Spreadsheet R	Select Google Sheets Account	nt	
Google Sheets New or Updated Spreadsheet Choose Account Dull in Samples	Google Sheets r@gmail.com @gmail.com added 2 months ago, used in 1 Zap	Test	
4 2. ACTION	Google Sheets added 33 minutes ago, used in 1 Zap	Test	
Set up this step	Connect an Account		
+ Add a Step		Save + Continue	

Figure 2.4: Choose the Google Docs account where the data is stored

5. Zapier provides you the list of the available **Google Spreadsheets** in your account so you can select the one you want to use. In our case, we are using our patient data saved in a Google Spreadsheet that looks like this:

ŝ		100% - \$	% .0 <u>_</u> .0 <u>0</u> 1	23 - Arial	~ 10 ~	BIŞ A		· ≡ · ± ·	÷ - ₽
fx									
	A	в	С	D	E	F	G	н	1
1	Glucose	Blood pressure	Insulin	BMI	Skinfold	Diabetes pedigree	Age	Pregnancies	
2	148	72	0	26.6	35	1.345	53	0	
3	85	66	0	26.6	35	1.345	53	0	
ŧ.	183	64	0	26.6	35	1.345	53	0	
5	89	66	94	26.6	35	1.345	53	0	
6	137	40	168	26.6	35	1.345	53	0	
7	116	74	0	26.6	35	1.345	53	0	
3	78	50	88	26.6	35	1.345	53	0	
9	115	0	0	26.6	35	1.345	53	0	
0	93	124	186	26.6	35	1.345	53	0	
1	93	124	186	26.6	35	1.345	53	0	
2	108	97	168	26.6	35	1.345	53	0	
3	108	97	168	26.6	35	1.345	53	0	
4									
5									
6									
7 B									
9									
0									
1									

Figure 2.5: Patient data in a Google Spreadsheet to be used as input to predict the likelihood of Diabetes

It is very important that the data you use to predict is compatible with your model data, i.e., the variables to predict should be the same as the ones used as inputs to train the the model. For this example, we previously trained a Machine Learning model in BigML using the Diabetes dataset¹ which contains the same variables found in the Google Spreadsheet to calculate the prediction (see Figure 2.5).

We can find our "Diabetes data" file in the Spreadsheet selector in Zapier as shown in Figure	2.6.
---	------

Dashboard \rightarrow Set Up Trigger Options or View Task History		YOUR ZAP IS	OFF
Name your zap Add a note	Set up Google Sheets Spreadsheet Row		
New or Updated Spreadsheet R	Spreadsheet (required)		
Google Sheets		~	
Y New or Updated Spreadsheet	Q Diabe		
Google Sheets teresa.asoler	Diabetes 1vL01PZSN-fs0voR4NuI-S_EgYK5pXZ8KH06NuWegxX4		
E Set Up Options			
Pull In Samples	Diabetes Prediction 122k1FdOMMo-rtLRiXbvW6Kf6YRa5C2JwWII3WoTTWDA		
+	⑦ Check Google Sheets & reload to bring in new choices.		
2. ACTION	Other Options		
2. ACTION	道 Use a Custom Value (advanced)		
Set up this step	Clear Current Choice		
+ Add a Step			
T Add a Step		Continue	
Get Help Response Time: ~2h M-F 9am-5pm PST			

Figure 2.6: Choose the Google Spreadsheet to use as a trigger

¹https://bigml.com/user/bigml/gallery/dataset/4f89c693155268645900003a

6. Then you need to select the specific **worksheet** where your data is stored (in this case it is stored in the Sheet 1) and also the **column** to trigger the action (you can choose any column or a specific column). Then click Continue.

Dashboard → Set Up Trigger Options or View Task Histor	Ŋ	*	YOUR ZAP IS	OFF
∕ Name your zap Add a note				
1. TRIGGER ··· New or Updated Spreadsheet R		Set up Google Sheets Spreadsheet Row		
Google Sheets		Spreadsheet (required) Diabetes	~	
Yew or Updated Spreadsheet Google Sheets teresa.asoler Edit Options		Worksheet (required) You must have column headers		
A Pull In Samples		Sheet1	×)	
* 2. ACTION		Trigger Column (optional) Trigger on changes to cells in this column only. Leave this field blank if you want the Zap to trigger on changes to an any, column	ny cell.	
Set up this step		© Refresh Fields		
+ Add a Step		Com	tinue	
Get Help Response Time: ~2h M-F 9am-5pm PST				

Figure 2.7: Choose the worksheet and the column to trigger the action

7. As a final step to set up your trigger, click the Fetch & Continue button. Zapier will attempt to access your document and test if it can fetch the data.

Dashboard \rightarrow Pull In Samples or <u>View Task History</u>	*	YOUR ZAP IS
/ Name your zap dd a note	Test Google Sheets	
New or Updated Spreadsheet R	THINGS TO DOUBLE-CHECK:	
Coogle Sheets New or Updated Spreadsheet Google Sheets teresa.asoler Edit Options	 Make sure you have at least one recent spreadsheet row created. Make sure it matches the trigger options you picked below: 	
A Pull In Samples	Spreadsheet Diabetes	
* 2. ACTION	Worksheet Sheet1	
4 Set up this step	Trigger Column any_column	
+ Add a Step		Fetch & Continue

Figure 2.8: Testing your trigger

8. If the test succeeds, click Continue to jump to the next step.

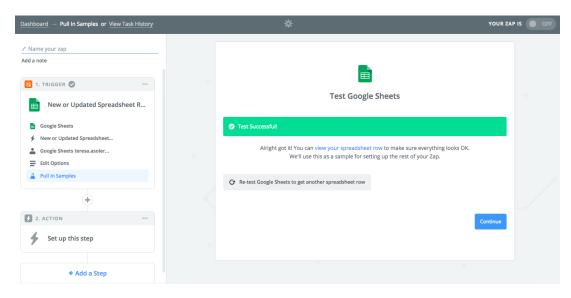


Figure 2.9: Finishing your trigger setup

2.2 Setting Up the Action: BigML Zapier App

In this section, we are going to instruct our Zap to use the new data from the previous step (see Section 2.1) to predict the likelihood of diabetes using an existing BigML model.

1. Search for the BigML Zapier app, as shown in the picture below.

Choose an Action A	φp
BigML	
	~
YOUR APPS	
show all	
	ETTA DigML

Figure 2.10: Select BigML Zapier app

2. Select **Create Prediction** from the list of available actions, then click **Save + Continue**. Find an explanation for the rest of available actions in Chapter 3.

$\underline{ \text{Dashboard}} \ \rightarrow \ \textbf{Choose Action or } \underline{ \text{View Task History}}$		*	YOUR ZA	P IS OFF
✓ Name your zap Add a note	Search Big	ML Actions		
	CREATE			
New or Updated Spreadsheet R	•	Create Prediction Predict using a model, logistic regression, or deepnets.		
2. ACTION	0	Create Centroid Find out the closest cluster to your data instance.		
O Set up this step	0	Create Anomaly Score Calculates the anomaly score of a data instance.		
BigML	-	Create Topic Distribution		
Choose Action Edit Template	0	Calculate all topic probabilities for a given document.		
Test this Step	SEARCH			
+ Add a Step	0	Find a Resource Finds a resource.		
Get Help Response Time: ~2h M-F 9am-5pm PST			Save + Continue	

Figure 2.11: Select Create Prediction out of the actions provided by the BigML Zapier app

- 3. Connect your BigML account to Zapier. For this, you need to follow these steps:
 - Connect an Account
- Click Connect an Account :

Figure 2.12: Connect your BigML account

• A modal window will be displayed asking you for your username and your API key:

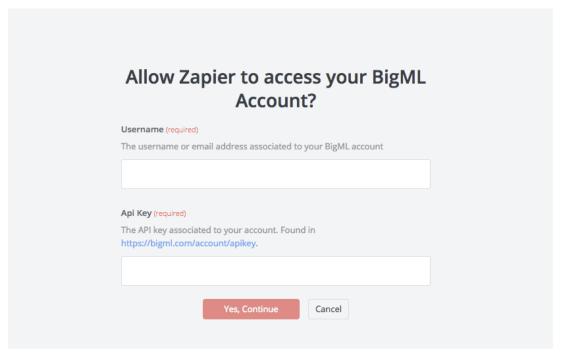


Figure 2.13: Set your BigML credentials

• You can find your API key in your BigML account² as shown in Figure 2.14:

PRODUCT • G	ETTING STARTED PRICING ▼ SUPPORT	ALFREDJR Dashbox
Plans & Pricing Organizations Certifications API Key	API Key Keep it secret!	G
Cloud Storages Payments User Account Info Email Settings Password Spread the Word Delete Account	Alternative Keys Shared Resources	Create New API Key

Figure 2.14: Find your BigML API key

• After you set your username and API key, you will be able to see your account connected in your Zap. Then click Save + Continue.

²https://bigml.com/account/apikey

$\underline{\text{Dashboard}} \ \rightarrow \ \text{Choose Account} \ or \ \underline{\text{View Task History}}$	* *	YOUR ZAP IS
/ Name your zap Add a note		
I. TRIGGER ··· Image: New or Updated Spreadsheet R	Select BigML Account	
÷	BigML (1.0.4) alfredjr alfredjr difa do days ago, used in 0 Zaps Test	
Create Prediction	Connect an Account	
 BigML Create Prediction 		
Choose Account Set Up Template	Save + Cont	inde
🔒 Test this Step		
+ Add a Step		

Figure 2.15: Your BigML account is connected to your Zap

4. Select which type of model you want to use for the prediction. In this case we are using an ensemble:

$Dashboard \ \rightarrow \ Set Up Create Prediction Template \ or \ \underline{View} Task History$	* You	UR ZAP IS OFF
Z Name your zap	0	
	BETA	
New or Updated Spreadsheet R	Set up BigML Prediction	
* · · · · · · · · · · · · · · · · · · ·	Setup Preview Learn more	9
Create Prediction	Resource Type (optional)	
BigML	Ensemble	×/
Create Prediction	Project Name (optional)	
BigML (1.0.4) alfredjr	Search a resource in this project (partial name accepted).	
🚍 Set Up Template		0
Test this Step		
	Resource Name (optional)	
	Search a resource with this name (partial name accepted).	
+ Add a Step		0
Get Help Response Time: ~2h M-F 9am-5pm PST		

Figure 2.16: Select the resource type to make predictions

5. Then you need to select the ensemble you want to use. In the "**Resource**" selector, you can select an ensemble out of the last 20 ensembles built in your BigML account or you can directly paste the ensemble ID (see Figure 2.17). If your ensemble is not found within the last 20 ensembles created, you can use the "Project Name", the "Resource Name" and/or the "Resource Tag" input boxes shown above to filter the models shown in this selector.

Dashboard Set Up Create Prediction Template or View Task History	* *	OUR ZAP I	S OFF
/ Name your zap Add a note	Project Name (optional) Search a resource in this project (partial name accepted).		
🔞 1. TRIGGER 🥝		Ξo	
New or Updated Spreadsheet R	Resource Name (optional) Search a resource with this name (partial name accepted).		
•		Ξo	
2. ACTION ··· Create Prediction	Resource Tag (optional) Search a resource with this tag.	Ξo	
BigML Create Prediction BigML (1.0.4) alfredjr	Resource (required) The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource ID from the resource URI show in your browser address bar. You can filter the listed resources using the filter fields above.		
Edit Template	Diabetes Diagnosis (bootstrap decision forest, 512-node, 10-model, deterministic order)	~	
	Pregnancies (optional) 1.0	_	
+ Add a Step		Ξo	
Get Help Response Time: ~2h M-F 9am-5pm PST	Glucose (optional) 1.0		

Figure 2.17: Select the ensemble to make predictions

Note: you need to select a model compatible with the data stored in the input spreadsheet selected in the first trigger step (see Section 2.1), i.e., the model needs to use the same variables stored in the Google Spreadsheet as inputs.

6. When you select your model, you need to define which input fields you want to use for the prediction. You can see in Figure 2.18 that for each of the model input fields, we are selecting a column in our Google Spreadsheet containing the patient data.

$\underline{\text{Dashboard}} \ \rightarrow \ \text{Set Up Create Prediction Template} \ \text{or} \ \underline{\text{View Task History}}$		YOUR ZAP IS
/ Name your zap	Exep 1 9	Ξο
Add a note	Glucose (optional) 1.0	
💽 1. TRIGGER 🕏 🚥	E Step 1 108	Ξo
New or Updated Spreadsheet R	Blood Pressure (optional) 1.0	
· · · · · · · · · · · · · · · · · · ·	(Step 1 97	Ξo
2. ACTION	Skinfold (optional) 1.0	
Create Prediction	5tep 1 28	Ξο
	Insulin (optional) 1.0	
BigML Create Prediction		Ξo
BigML (1.0.4) alfredjr	Blood pressure 97	
🚍 Edit Template	Pregnancies 9	
Test this Step	Skinfold 28	
	Insulin 168	
+ Add a Step	BMI 27	

Figure 2.18: Inputs for the prediction

7. You can configure many other parameters to make your prediction, all of which are explained in Chapter 3. However, for this example, we use all the default parameter values and click Continue.

$\underline{Dashboard} \ \rightarrow \ \textbf{Set Up Create Prediction Template} \ or \ \underline{View Task History}$		YOUR ZAP IS
✓ Name your zap	Last prediction	~
Add a note	Operating Kind (optional)	
🔞 1. TRIGGER 🥏	Specifies the method that should be used to combine predictions in an ensemble.	
	Probability	\sim +
New or Updated Spreadsheet R		
	Positive Class (Classif. Only) (optional) When present, BigML will predict the positive class if its probability, confidence or votes (depending on the oper	ating kind)
	is greater than the threshold set below. Otherwise, it will predict the highest-scoring class.	and and
2. ACTION		Ξo
Create Prediction	Threshold (Classific. Only) (optional) 1.0	
BigML	When the positive class is present, this is the threshold used to predict it. A number between 0 and 1.	
Create Prediction	0	Ξo
BigML (1.0.4) alfredjr		
🚍 Edit Template	Of Refresh Fields	
A Test this Step		
		ontinue
+ Add a Step		
Cat Help Personna Timer -2h I M E Cam Fam DCT		

Figure 2.19: Click "Continue" when you finish setting the prediction parameters

8. At this point you can run a test by clicking Send Test to BigML to check that everything works. This tests will take the last row in your Google Spreadsheet and it will use the selected BigML model to make a prediction.

$\underline{\text{Dashboard}} \ \rightarrow \ \text{Test this Step or } \underline{\text{View Task History}}$		* *	UR ZAP	IS OFF
✓ Name your zap		Additional data: Missing Strategy:		
I. TRIGGER ♥ ···		Operating Kind:		
New or Updated Spreadsheet R		Positive Class (Clas Threshold (Classifi		
* 2. ACTION		EMPTY FIELDS:		
Create Prediction		Project Name: empty (optional) Resource Name: empty (optional)		
 BigML Create Prediction 		Resource Tag: empty (optional) Additional data: empty (optional) Positive Class (Clas empty (optional)		
 BigML (1.0.4) alfredjr Edit Template Test this Step 				
+ Add a Step	S	Skip Test	J	

Figure 2.20: Send a test to check the Zap works

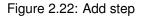
If the test succeeds, you will find a new prediction in your BigML account and in the Zap result (see the predictions for the classes "True" and "False" in Figure 2.21)

Dashboard \rightarrow Test this Step or View Task History	*	YOUR ZAP IS OFF
/ Name your zap Add a note	Q Search	
🔞 1. TRIGGER	confidences: 0:	
New or Updated Spreadsheet R	u: 1: False 2: 0.75592 1: 1: True	
Create Prediction	2: 0.1726 error_predictions: 0	
BigML Create Prediction	code: 201	
Greater Frediction BigML (1.0.4) alfredjr Edit Template	missing_strategy: 0	
Test this Step	Re-test This Step	Add a Step or Finish
+ Add a Step		

Figure 2.21: See the prediction result

This may be the end of the workflow if we just want the new prediction to be stored in our BigML account. However, since we want to add a third step to send this prediction via e-mail, we need to click Add a Step .

$\underline{Dashboard} \ \rightarrow \ \textbf{Test this Step or} \ \underline{View Task \ History}$	*	YOUR ZAP	IS OFF
/ Name your zap	Q Search		
Add a note	confidences:		
New or Updated Spreadsheet R	0: 1: False		
•	2: 0.75592		
5 2. ACTION 🕲	1: True		
Create Prediction	2: 0.1726 error_predictions: 0		
BigML Create Prediction	code: 201		
BigML (1.0.4) alfredjr Edit Template	missing_strategy: 0		
Test this Step	Re-test This Step Add a Step o	Finish	
+ Add a Step			
Get Hain Deconnee Timer ~7h M.E Gam. Enm DCT			



2.3 Sending an E-Mail with the Prediction: Gmail App

In this section, we are going to instruct our Zap to send the prediction generated in our previous step (see Section 2.2) via e-mail.

1. As you did before, select the Gmail app in Zapier.

Dashboard \rightarrow Choose App or View Task History	* You	R ZAP IS OFF
/ Name your zap Add a note	Choose an Action App	
Create Prediction BigML (1.0.4) alfredjr	\$earch hundreds of apps	,
Edit Template	Google Sheets	1 +
Test this Step	Gmail	
◆ 3. ACTION ····	Slack	
Set up this step	KailChimp	
/ Choose App	Google Calendar	
Choose Action	Trello	
 Set Up Template Test this Step 	y Twitter	
	Facebook Lead Ads	
Add a Step Get Help Response Time: -2h M-F 9am-Spm PST	Narrow your search to reveal 1116 hidden results.	

Figure 2.23: Select the Gmail app

2. Select the Send Email option and then click Save+Continue.

$\underline{ \text{Dashboard}} \ \rightarrow \ \textbf{Choose Action or } \underline{ \text{View Task History}}$	*	OUR ZAP IS
Name your zap Add a note Create Prediction BigML (1.0.4) alfredjr	Select Gmail Action	
Edit Template	Search Gmail Actions	
* 3. ACTION ***	CREATE Create Draft Create (but do not send) a new email message.	
Set up this step	• Send Email Create and send a new email message.	
Choose Action Edit Template	show less common options SEARCH	
Test this Step	 Find Email Finds an email message. Optionally, create one if none are found. 	
+ Add a Step	Save + Cont	inue

Figure 2.24: Select the Send Email option

3. Connect the Zap to your Gmail account.

$\underline{ashboard} \ \rightarrow \ \mathbf{Choose} \ \mathbf{Account} \ \ \mathbf{or} \ \ \underline{View} \ \mathbf{Task} \ History$	*	YOUR ZAP IS
Name your zap		
d a note	M	
Create Prediction		
BigML (1.0.4) alfredjr	Select Gmail A	Account
Edit Template		
A Test this Step	Connect an Account	
+		
3. ACTION		
Send Email		Continue
M Gmail		
🗲 Send Email		
Choose Account		
Set Up Template		
Caracteris Step		
+ Add a Step		
Get Help Response Time: ~2h M-F 9am-5pm PST		

Figure 2.25: Connect your Gmail account

$\underline{\text{Dashboard}} \rightarrow \text{Choose Account or } \underline{\text{View Task History}}$	*	YOUR ZAP IS OFF
	Select Gmail Account	
Test this Step	Gmail added just now, used in 0 Zaps Connect an Account	Test
Gmail Send Email Choose Account	Sav	e + Continue
Choose Account Set Up Template Test this Step		
+ Add a Sten Get Help Response Time: -2h M-F 9am-5pm PST		

When it is connected, click Save+Continue .

Figure 2.26: Click Save + Continue

4. You need to configure the subject and the body of the e-mail, and you can optionally configure other parts like the adress to send the e-mail.

$\underline{\text{Dashboard}} \rightarrow \text{Set Up Send Email Template or } \underline{\text{View Task H}}$	listory	* YOUR ZA	P IS OFF
Name your zap Add a note Create Prediction BigML (1.0.4) alfredjr Edit Template		Set up Gmail Email	
Test this Step		Setup Preview Learn.more No	
		+ M Cc (optional) Who should be ccd on this email? Click the + button to add another email address field, or separate multiple email addresses in one field with a comma eg. test€example.com, example€example.com.	

Figure 2.27: Configure the e-mail

For the body, you can select the item from the second step (Section 2.2) that includes the predicted classes probabilities as shown in Figure 2.28.

Dashboard \rightarrow Set Up Send Email Template or View Task History	*	YOUR ZAP IS
Name your zap Add a note ✓ Create Prediction BigML (1.0.4) alfredjr Edit Template Test this Step	M Body Type (optional) If using the HTML option, you must add any and all formatting (paragraphs, lists, etc) directly using HTML Lean Plain M Body (required) @ Stop 21 ["False",0.79381]; ["True",0.20819]	50 more.
Action ···· Send Email	M Label/Malibox (optional)	-
Send Email Ginal alvarez@bigml.com Edit Template Test this Step Get Help Response Time: -2h M=F 9am-Spm PST	M Attachments (optional) A file to be attached. Can be an actual file or a public URL which will be downloaded and attached.	-

Figure 2.28: Configure the e-mail

5. When you finish the e-mail configuration, you can click the Send Test to Gmail button. If the Zap works correctly, the e-mail containing the predicted class probabilities should be sent.

$\underline{\text{Dashboard}} \ \rightarrow \ \text{Test this Step or} \ \underline{\text{View Task History}}$			YOUR ZAP	P IS OFF
Name your zap Add a note Create Prediction BigML (10.4) alfredjr Edit Template	Body Type: Body: Label/Mailbox: Attachments:	plain False,0.79381,True,0.20619		
 Test this Step Test this Step ACTION O Send Email Send Email Send Email Gmail alvarez@bigml.com Edit Template Test this Step 	Cc: Bcc: From: From Name: Reply To: Label/Mallbox: Attachments:	empty (optional) empty (optional) empty (optional) empty (optional) empty (optional) empty (optional)	Small	
+ Arid a Sten Get Help Response Time: -2h M-F 9am-Spm PST				

Figure 2.29: Send a test

6. If the test succeeds, click Finish .

$\underline{\text{Dashboard}} \rightarrow \text{Test this Step or } \underline{\text{View Task History}}$		YOUR ZAP	IS OFF
✓ Name your zap Add a note	A Test email was sent to Gmail about 11 seconds ago.		
🔞 1. TRIGGER 🖉 🚥			
New Resource	SAMPLE FROM GMAIL: Q. Search		*
2. ACTION 🕲 …	labelids:		
Send Email	2: SENT 3: INBOX		
M Gmail Send Email	ld: 163a82c4a1e91691 threadid: 163a82c4a1e91691		
Grail alvarez@bigml.com	uireauu. 1038020481531031		
A Test this Step	Re-test This Step Or Fin	ish	
+ Add a Step			
Get Help Response Time: ~2h M-F 9am-5pm PST			

Figure 2.30: Finishing your Zap creation

2.4 Turning Your Zap On

Finally, you will have the chance to switch your Zap on so it is executed whenever new data is present in your trigger (2.1) or at a fixed time interval. While on, your Zap will automatically calculate predictions based on new data coming in (2.2) and send them via e-mail (2.3).

Dashboard or View Task History		YOUR ZAP IS OFF
✓ Name your zap Add a note ■ BigML (1.0.4) alfredjr ≡ Edit Template ⊒ Test this Step	Ready to turn on your Zap? We recommend giving your Zap a name.	
Send Email Send Email Send Email	VOUR ZAP IS OFF While on, this Zap will automagically check for your Google Sheets New Spreadsheet Row every © 15 minutes.	
Graal alvarez@bigml.com Edit Template Crest this Step Add a Step		
Get Help Response Time: ~2h M-F 9am-5pm PST		

Figure 2.31: Turn your Zap on



BigML Zapier App Actions

This chapter details all the actions that you can select using the BigML Zapier app. These actions can be classified in two groups as you can see in Figure 3.1:

- **CREATE**: you can select a supervised or unsupervised model and make predictions using the following actions:
 - Create Prediction: make predictions using a model, ensemble, logistic regression, or deepnet. See Section 3.2.
 - Create Centroid: make cluster predictions. See Section 3.3.
 - Create Anomaly Score: make anomaly predictions. See Section 3.4.
 - Create Topic Distribution: make topic model predictions. See Section 3.5.
- SEARCH: instead of selecting a specific model to be used to make predictions, you can dynamically select it based on a number of possible criteria, such as the name, project, and tags using the **Find a Resource** action (see Section 3.6). The selected model can be then "plugged" into the workflow to make predictions. This option is very useful if you frequently retrain your model with new data.

$\underline{\text{Dashboard}} \ \rightarrow \ \textbf{Choose Action or} \ \underline{\text{View Task History}}$		* You	JR ZAP IS	OFF
✓ Name your zap Add a note	Search Big	ML Actions		
	CREATE			
New or Updated Spreadsheet R	•	Create Prediction Predict using a model, logistic regression, or deepnets.		
* 2. ACTION	0	Create Centroid Find out the closest cluster to your data instance.		
Set up this step	0	Create Anomaly Score Calculates the anomaly score of a data instance.		
BigML Choose Action Choose Ac	0	Create Topic Distribution Calculate all topic probabilities for a given document.		
Test this Step	SEARCH			
	0	Find a Resource Finds a resource.		
+ Add a Step				
Get Help Response Time: ~2h M-F 9am-5pm PST		Save + Continu	e	

Figure 3.1: Actions provided by the BigML Zapier app

To set up any of these actions in a Zap, you need to specify a series of **arguments** such as the BigML resource you want to use and the missing strategy for your predictions. The arguments related to the

resource selection are the same for all the CREATE and SEARCH actions, and each action can also have other specific arguments. The Section 3.1 describes **common arguments for all actions** while the following sections (from Section 3.2 to Section 3.5) detail **arguments specific** to each action.

3.1 Set Up Common Arguments for All Actions

To make the BigML Zapier app work, you need to set up some **basic arguments** that are **common to all the app actions** previously described: the Resource Type, Project Name, Resource Name, Resource Tag and Resource as shown in Figure 3.2.

Set up BigML Centroid
rn more
\sim
artial name accepted).
Ξ
partial name accepted).
Ξ
Ξ
=
edict e.g. ensemble/12344567. You can get the resource ID from the resource URL as r. You can filter the listed resources using the filter fields above.
·

Figure 3.2: Common arguments for all actions

All these arguments are related to the way you **select the resource** to be used either for the prediction or the search actions. The **Resource** argument is the only one required; the rest of them act as filters to find the resource you want:

Resource Type (optional): for the Create Prediction action you can choose a model, an ensemble, a logistic regression or a deepnet. For the Create Centroid, Create Anomaly Score, or Create Topic Distribution, you can find a pre-selected cluster, anomaly detector, or topic model respectively and you cannot select other type of model or the prediction will fail. The Find a Re-

source action allows you to select any of the mentioned resources and additionally, a WhizzML script, a prediction, or an OptiML.

- Project Name (optional): if your resource is found in a project in your BigML Dashboard, you can
 include the name here, so the resources listed in the Resource selector (below) will be filtered by
 the given project.
- **Resource Name** (optional): you can include the resource name (or part of the resource name) so the resources listed in the **Resource** selector (below) will be filtered by the given name.
- **Resource Tag** (optional): if your resource is tagged, you can include any of the tags here so the resources listed in the **Resource** selector (below) will be filtered by the given tag.
- **Resource** (required): the specific model you want to use for the prediction. You can either choose the resource using the selector that shows the last 20 created resources filtered by the **Resource Type**, the **Project Name**, the **Resource Name**, and the **Resource Tag** explained above.

For example, in the Figure 3.3 you can see the list of the last 20 created resources displayed in the **Resource** selector.

Search a resource with this name (partial name accepted).	
	Ξ
Resource Tag (optional)	
Search a resource with this tag.	
	Ξ
	_
The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource	
The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource	
The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource	s above.
 Resource (required) The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource shown in your browser address bar. You can filter the listed resources using the filter field Search Diabetes diagnosis [auto-scaled fields, critical value=5] cluster/5b0c8176eba31 	s above.
The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource shown in your browser address bar. You can filter the listed resources using the filter field	s above.

Figure 3.3: Display the last 20 created resources

You can filter that list by a resource name by typing part of the name (see Figure 3.4).

Loan	=
LOAN	
Resource Tag (optional)	
Search a resource with this tag.	
	Ξ
Resource (required) The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource ID fron shown in your browser address bar. You can filter the listed resources using the filter fields above	2.
The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource ID from shown in your browser address bar. You can filter the listed resources using the filter fields above	
The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource ID from shown in your browser address bar. You can filter the listed resources using the filter fields above	e.
The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource ID from shown in your browser address bar. You can filter the listed resources using the filter fields above	e.
The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource ID from shown in your browser address bar. You can filter the listed resources using the filter fields above	6d

Figure 3.4: Filter resources by name

Alternatively, instead of selecting a resource from the selector, you can select the option **Use Custom Value (advanced)** (see Figure 3.5) and then paste the full identifier of the resource to be used, e.g., model/1234567890ABCEF (see Figure 3.6). By using this option, you will need to click on the Refresh Fields button to display the input data for the selected resource (explained below).

Dashboard \rightarrow Set Up Create Centroid Template or View Task History	*	YOUR ZAP	IS OFF
∕ Name your zap Add a note	Search a resource with this name (partial name accepted).	Ξo	
	Resource Tag (optional) Search a resource with this tag.		
New Spreadsheet Row	Resource (required) The ID of the resource to use to predict e.g. ensemble/12344567, You can get the resource ID from the resource	URL as	
2. ACTION ··· Create Centroid	shown in your browser address bar. You can filter the listed resources using the filter fields above.	~	
BigML (1.0.5) Create Centroid BigML (1.0.5) teresa3	hr (2) dataset filte [auto-scaled fields, critical value=5] cluster/5852755c01440443bb00040d hr (2) dataset filte [auto-scaled fields, critical value=5] cluster/5852752201440443b7000a0e Customer Churn datas [auto-scaled fields, critical value=5] cluster/5847e65749c4a16ae5001ffd		
E Set Up Template	Check BigML (1.0.5) & reload to bring in new choices. Other Options	^	
+ Add a Step	^(그) Use a Custom Value (advanced)		

Figure 3.5: Select Custom Value option

Dashboard		YOUR ZAP IS	OFF
✓ Name your zap	Search a resource with this name (partial name accepted).	Ξo	
💽 1. TRIGGER 🧭 🚥	Resource Tag (optional) Search a resource with this tag.		
New Spreadsheet Row		Ξo	
* 2. ACTION ····	Resource (required) The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource ID from the resource shown in your browser address bar. You can filter the listed resources using the filter fields above.	URL as	
Create Centroid	Use a Custom Value (advanced)	~	
 BigML (1.0.5) Create Centroid 	Custom Value for Resource ID (required) Provide the Resource ID, not the Name here. Not sure how to use custom values? Learn more here.		
BigML (1.0.5) teresa3	cluster/5b0c8176eba31d2b44000298	Ξo	
Edit Template	Ø Refresh Fields		
+ Add a Step	c	ontinue	

Figure 3.6: Select Custom Value option

Note: you do not neet to specify a Resource for the Find a Resource action since this action dynamically searches for different resources by all the other properties (see Section 3.6).

When you select the resource you want to use for all the **PREDICT actions**, a list of **the first five input fields in the model** will be displayed as new arguments so they can be **mapped to the input data** you want to use for the prediction. If you use the Custom Value option, you will need to click on the Refresh Fields button to display the input data of the resource selected:

• Input fields (optional): the first five input fields used by the model selected will be displayed. You can map each of these fields to the input data you want to use for the prediction. If you do not map a given field, it will not be used for the prediction (it will count as a missing value for the BigML model). Clusters do not allow missing values so you need to provide all the input fields that the cluster uses to make the prediction.

If you previously connect the first step in the Zap to a Google Spreadsheet or any other application, you can map your fields to the exact columns where the data to be used as an input for the prediction is stored (see Figure 3.7).

Resource (required)	
The ID of the resource to use to predict e.g. ensemble/1234 shown in your browser address bar. You can filter the listed	4567. You can get the resource ID from the resource URL as resources using the filter fields above.
Diabetes diagnosis [auto-scaled fields, critical value=5	· · · · · · · · · · · · · · · · · · ·
Pregnancies (optional) 1.0	
 B Step 1 No data Plasma Glucose (optional) 1.0 	
Plasma Glucose (optional) 1.0	
 Plasma Glucose (optional) 1.0 glu New Spreadsheet Row Glucose 	Ē
 Plasma Glucose (optional) 1.0 glul New Spreadsheet Row 	

Figure 3.7: Map the model fields with the input data

• Number of fields to display (optional): if your model contains more than five fields and you want to use them for the prediction, you can choose a higher number of fields to display using this option.

Number of input fields t	o display (optional) 1.0		
he number of input fields to			
5			\sim
Additional data (optional))		
ou can provide here input f ase sensitive (i.e., Count is n		not listed above. Please, keep in mind field names	s are
		Ξō	-
			+
🗘 Refresh Fields			

Figure 3.8: Select the numebr of fields to display

• Additional data (optional): you can also input single additional fields (used by the model) and map

Step 1 No data			Ξc
Number of input fields to	o display (optional) 1.0		
he number of input fields to	show.		
5			~
ase sensitive (i.e., Count is no	ot the same as count).		
			=o -
🗘 Refresh Fields			Ξο - +

them to the input data to be used in the prediction.

Figure 3.9: Include additional fields to map

3.2 Create Prediction

The **Create Prediction** action allows you to create a prediction using an existing classification or regression model, ensemble, logistic regression, or deepnet.

Dashboard \rightarrow Choose Action or View Task History		*	YOUR ZAP IS	OFF
/ Name your zap Add a note		ML Actions		
	CREATE			
New or Updated Spreadsheet R	•	Create Prediction Predict using a model, logistic regression, or deepnets.		
(+) 2. ACTION	0	Create Centroid Find out the closest cluster to your data instance.		
O Set up this step	0	Create Anomaly Score Calculates the anomaly score of a data instance.		
BigML Choose Action Edit Template	0	Create Topic Distribution Calculate all topic probabilities for a given document.		
Control Test this Step	SEARCH			
	0	Find a Resource Finds a resource.		
+ Add a Step				
Get Help Response Time: ~2h M-F 9am-5pm PST			Save + Continue	

Figure 3.10: Create Prediction

To set up a prediction using any of the aforementioned models, you need to select the resource and define the input fields you want to use by providing the arguments explained in Section 3.1. Moreover, you can optionally set the following arguments to configure the way predictions are calculated:

- **Missing Strategy** (optional): you can choose the strategy used to handle missing values. This can be either "Last Prediction" (by default) or "Proportional".
- **Operating Kind** (optional): you can choose the strategy used to calculate the predictions: probability, confidence or votes (only for non-boosted ensembles). The operating kind only affects the results for ensembles and models; for the other resources, confidence and probability will always yield the same results.
- **Positive Class (Classif. Only)** (optional): when using a **Threshold**, BigML only predicts the positive class if its probability, confidence or votes (depending on the operating kind) is greater than the threshold set. Otherwise, it will predict the following highest-scoring class. If no class is specified, the category that appears least frequently in the training data is chosen.
- **Threshold (Classif. Only)** (optional): the positive class will be predicted only if it is present, and the probability, confidence or votes (depending on the operating kind) is greater than the threshold set. The threshold needs to be a number between 0 and 1.

For a more detailed explanation of all the above arguments, see the document Classification and Regression with the BigML Dashboard. [1].

Specifies the method that should be used when a missing valu are:	ue is found in the input data for a decision node. The optior
- last prediction predicts based on the subset of the data w	hich reached the parent of the missing split.
 proportional evaluates all the subtrees of a missing split a lata in each subtree. 	nd recombines their predictions based on the proportion o
Last prediction	~
Operating Kind (optional)	
Specifies the method that should be used to combine predicti	ons in an ensemble.
Probability Positive Class (Classif. Only) (optional) When present, BigML will predict the positive class if its proba	bility, confidence or votes (depending on the operating kind
Probability Positive Class (Classif. Only) (optional) When present, BigML will predict the positive class if its proba	bility, confidence or votes (depending on the operating kind
	bility, confidence or votes (depending on the operating kind
Probability Positive Class (Classif. Only) (optional) When present, BigML will predict the positive class if its proba s greater than the threshold set below. Otherwise, it will pred	bility, confidence or votes (depending on the operating kind lict the highest-scoring class.
Probability Positive Class (Classif. Only) (optional) When present, BigML will predict the positive class if its proba s greater than the threshold set below. Otherwise, it will pred Threshold (Classific. Only) (optional) 1.0	bility, confidence or votes (depending on the operating kind ict the highest-scoring class.
Probability Positive Class (Classif. Only) (optional) When present, BigML will predict the positive class if its proba s greater than the threshold set below. Otherwise, it will pred Threshold (Classific. Only) (optional) 1.0	bility, confidence or votes (depending on the operating kind ict the highest-scoring class.
Probability Positive Class (Classif. Only) (optional) When present, BigML will predict the positive class if its proba is greater than the threshold set below. Otherwise, it will pred Threshold (Classific. Only) (optional) 1.0 When the positive class is present, this is the threshold used t	bility, confidence or votes (depending on the operating kind ict the highest-scoring class.

Figure 3.11: Set up BigML prediction

3.3 Create Centroid

The **Create Centroid** action allows you to create a prediction using an existing cluster.

$\underline{\text{Dashboard}} \ \rightarrow \ \text{Choose Action or} \ \underline{\text{View Task History}}$			YOUR ZAP	
/ Name your zap	Search Big	(ML (1.0.5) Actions		
	0	Create Prediction Predict using a model, logistic regression, or deepnets.		
New Spreadsheet Row	0	Create Centroid Find out the closest cluster to your data instance.		
2. ACTION	0	Create Anomaly Score Calculates the anomaly score of a data instance.		
Create Prediction		show less common options		
BigML (1.0.5)	SEARCH			
f Create Prediction		Find a Resource		
BigML (1.0.5) alfredjr	0	Finds a resource.		
Edit Template				
Test this Step			Save + Continue	
•				
Cat Help Desponse Time: «2h M.E.Gam.Spm PST				

Figure 3.12: Create Centroid

To set up a centroid, you need to select the resource and define the input fields you want to use by providing the arguments explained in Section 3.1.

3.4 Create Anomaly Score

The Create Anomaly Score action allows you to create a prediction using an existing anomaly.

$\underline{\text{Dashboard}} \ \rightarrow \ \textbf{Choose Action or} \ \underline{\text{View Task History}}$		* YOUR	ZAP IS	OFF
/ Name your zap	CREATE	tur fi tanàh renormati		
🔞 1. TRIGGER 🖉	0	Create Prediction Predict using a model, logistic regression, or deepnets.		
New Spreadsheet Row	0	Create Centroid Find out the closest cluster to your data instance.		
• · · · · · · · · · · · · · · · · · · ·	•	Create Anomaly Score Calculates the anomaly score of a data instance.		
Create Centroid BigML (1.0.5)	0	Create Topic Distribution Calculate all topic probabilities for a given document.		
 BigML (1.0.5) Create Centroid 	SEARCH			
BigML (1.0.5) teresa3 Edit Template Test this Step	0	Find a Resource Finds a resource.		
		Save + Continue		
Get Help Response Time: ~2h M-F 9am-5pm PST				

Figure 3.13: Create Anomaly Score

To set up an anomaly score, you need to select the resource and define the input fields you want to use by providing the arguments explained in Section 3.1.

3.5 Create Topic Distibution

The Create Topic Distribution action allows you to create a prediction using an existing topic model.

$\underline{Dashboard} \to \text{Choose Action or } \underline{View Task History}$		* You	IR ZAP IS	••	DFF
Name your zap Add a note	CREATE				
💽 1. TRIGGER 🔮 🛛	0	Create Prediction Predict using a model, logistic regression, or deepnets.			
New Spreadsheet Row	0	Create Centroid Find out the closest cluster to your data instance.			
(+) 2. ACTION	0	Create Anomaly Score Calculates the anomaly score of a data instance.			
Create Centroid	•	Create Topic Distribution Calculate all topic probabilities for a given document.			
BigML (1.0.5) Create Centroid	SEARCH				
 ■ BigML (1.0.5) teresa3 ■ Edit Template ▲ Test this Step 	0	Find a Resource Finds a resource.			
+ 		Save + Continu	e		

Figure 3.14: Create Topic Distribution

To set up a topic distribution, you need to select the resource and define the input fields you want to use by providing the arguments explained in Section 3.1.

3.6 Find a Resource

The **Find a Resource** search allows you to specify a number of criteria to identify the resource you would like to feed into a subsequent prediction step. For example, you could use the search action to feed the prediction action with the latest version of a model whose name contains a given string, or the latest version of a model belonging to a given project, identified through its name or tags. This option is very useful; e.g. if you have several versions of the same model because you frequently retrain your model with new data.

Dashboard \rightarrow Choose Search or View Task History		*	YOUR ZAP I	S OF
✓ Name your zap Add a note	Search Big	zML (1.0.5) Actions		
	CREATE			
New Spreadsheet Row	0	Create Prediction Predict using a model, logistic regression, or deepnets.		
•	0	Create Centroid Find out the closest cluster to your data instance.		
2. SEARCH ···· Find a Resource	0	Create Anomaly Score Calculates the anomaly score of a data instance.		
BigML (1.0.5)		show less common options		
Find a Resource	SEARCH			
BigML (1.0.5) teresa3 Set Up Options	•	Find a Resource Finds a resource.		
Cest this Step				
+			Continue	

Figure 3.15: Find Resource

To set up the Find a Resource action, you need to configure the arguments explained in Section 3.1 to select the resource you want.

When you set up your search action, you can use a following prediction action to associate the **Re-source** argument of your prediction action with the result produced by the search action. This can be accomplished by clicking the **Custom Value (advanced)** option of the Resource argument as shown in Figure 3.16.

Dashboard \rightarrow Set Up Create Prediction Template or View Task History		YOUR ZAP IS
✓ Name your zap	ала стал сола се чита сталини филим. на не иссерсор-	Ξō
Q 2. SEARCH 🖉 ***	Resource Tag (optional) Search a resource with this tag.	
Find a Resource		Ξo
3. ACTION	Resource (required) The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource ID from the resour shown in your browser address bar. You can filter the listed resources using the filter fields above.	ce URL as
BigML (1.0.5) Create Prediction	Diabetes diagnosis [512-node, deterministic order] model/5b0d943508b07e46280022a4 Diabetes diagnosis [512-node, deterministic order] model/5b0d943508b07e46280022a2	
▲ BigML (1.0.5) teresa3 ➡ Set Up Template	Diabetes diagnosis [512-node, deterministic order] model/5b0d943508b07e46280022a0 Diabetes diagnosis [512-node, deterministic order] model/5b0d943508b07e462800229e	
A Test this Step	♂ Check BigML (1.0.5) & reload to bring in new choices.	
+ Add a Step	Other Options Use a Custom Value (advanced)	^
Get Help Response Time: -2h M-E 9am-Som PST	Clear Current Choice	_

Figure 3.16: Select the Custom Value option for the Resource argument

Then you need to select the **ID option** associated to the Find a Resource previous step as shown in Figure 3.17.

$\underline{\text{Dashboard}} \ \rightarrow \ \text{Set Up Create Prediction Template or } \underline{\text{View Task History}}$	*	YOUR ZAP	IS OFF
✓ Name your zap		Ξō	
Add a note	Resource (required)		
Q 2. SEARCH 🖉	The ID of the resource to use to predict e.g. ensemble/12344567. You can get the resource ID from the resource shown in your browser address bar. You can filter the listed resources using the filter fields above.	e URL as	
Find a Resource	Use a Custom Value (advanced)	~	
+	Custom Value for Resource ID (required)		
🛃 3. ACTION	Provide the Resource ID, not the Name here. Not sure how to use custom values? Learn more here.		
		Ξō	
Create Prediction	Q Search		
BigML (1.0.5)	1 🖥 New Spreadsheet Row	\sim	
Create Prediction BigML (1.0.5) teresa3	②	^	
🚍 Set Up Template	Name Diabetes diagnosis [[512-node, pruned, deterministic order]		
Test this Step	ID model/5b0dc0d12a83475c8b000a3a		
	Categorical true		
+ Add a Step	Resource		
Get Help Response Time: ~2h M-F 9am-5pm PST	Missing choices? Re-test step 2.		

Figure 3.17: Select the ID of the resource from the Find a Step action

At this point, you need to manually input the field names used by the selected resource and map them to the data you want to use for the prediction (see Figure 3.18 and Figure 3.19). When you finish mapping the fields, you can click Continue to finish the workflow.

$\underline{Dashboard} \ \rightarrow \ \text{Set Up Create Prediction Template or } \underline{View Task History}$	*		YOUR ZAP	IS OFF
✓ Name your zap	Provide the Resource ID, not the Name F	nere. Not sure how to use custom values? Learn more here.	Ξo	
I 2. SEARCH I I I I I I I I I I I I I I I I I I I	Additional data (optional) You can provide here input for any featu case sensitive (i.e., Count is not the same	ires of your model that is not listed above. Please, keep in mind fii a as count).	eld names are	
•	plasma glucose		Ξο -	
3. ACTION		Q Search		
Create Prediction	O Refresh Fields	Blood pressure 75		
 BigML (1.0.5) Create Prediction 		Insulin 145 BMI 26.6	Intinue	
BigML (1.0.5) teresa3 Edit Template		Skinfold 35	- /	
Test this Step		Diabetes pedigree 1.345		
+ Add a Step		Age 53 Glucose 120	-	
Get Help Response Time: ~2h M-F 9am-5pm PST		Pregnancies 0		

Figure 3.18: Use the Additional data argument to add the resource input fields names and select the corresponding data from a previous step

$\underline{\text{Dashboard}} \ \rightarrow \ \text{Set Up Create Prediction Template} \ \text{or} \ \underline{\text{View T}}$	ask History				YOUR ZAP IS	OFF
✓ Name your zap Add a note		Use a Custom Value (advance	ar. You can litter the listed resources using the litte	r tields above.	~	
Q 2. SEARCH 🛇		Custom Value for Resource ID (re				
Find a Resource		Provide the Resource ID, not the l	Name here. Not sure how to use custom values? Le	arn more here.	Ξο	
+		Additional data (optional) You can provide here input for an case sensitive (i.e., Count is not the case sensitive (i.e., Count is not the count of the count	ry features of your model that is not listed above. P ie same as count).	lease, keep in mind field name	is are	
Create Prediction		plasma glucose	Etep 1 120	Ξō	-	
BigML (1.0.5)		bmi	E Step 1 26.6	Ξō	-	
Create Prediction BigML (1.0.5) teresa3		age	53 Step 1	Ē	-	
🚍 Edit Template		blood presssure	🕞 Step 1 75	Ξo	-	
Test this Step					+	
+ Add a Step		O Refresh Fields				
Get Heln Resonnce Time: ~2h I M-F 9am-5nm PST						

Figure 3.19: Map the input fields of the selected resource to the input data from previous steps

List of Figures

2.1	Create a new Zap					3
2.2	Choose the Google Sheets app					3
2.3	Choose the Google Sheets app as the trigger					4
2.4	Choose the Google Docs account where the data is stored					4
2.5	Patient data in a Google Spreadsheet to be used as input to predict the like	əlih	1000	d c	of	
	Diabetes					5
2.6	Choose the Google Spreadsheet to use as a trigger					5
2.7	Choose the worksheet and the column to trigger the action					6
2.8	Testing your trigger					6
2.9	Finishing your trigger setup					7
2.10	Select BigML Zapier app					7
2.11	Select Create Prediction out of the actions provided by the BigML Zapier app					8
2.12	Connect your BigML account					8
	Set your BigML credentials					9
	Find your BigML API key					9
	Your BigML account is connected to your Zap					10
	Select the resource type to make predictions					10
2.17	Select the ensemble to make predictions					11
2.18	Inputs for the prediction	÷				11
2.19	Click "Continue" when you finish setting the prediction parameters	÷				12
2 20	Send a test to check the Zap works			• •	•	12
2.21	See the prediction result	÷				13
2 22	Add step	•	•••	•••	·	13
	Select the Gmail app					14
	Select the Send Email option					14
	Connect your Gmail account					15
	Click Save + Continue					15
	Configure the e-mail					16
2 28	Configure the e-mail	·	•••	• •	•	16
	Send a test					17
	Finishing your Zap creation					17
2.31	Turn your Zap on	·	•••	• •	•	18
2.01		·	• •	• •	•	10
3.1	Actions provided by the BigML Zapier app					19
3.2	Common arguments for all actions					21
3.3	Display the last 20 created resources					22
3.4	Filter resources by name					23
3.5	Select Custom Value option					23
3.6	Select Custom Value option				÷	24
3.7	Map the model fields with the input data					25
3.8	Select the numebr of fields to display					25
3.9	Include additional fields to map					26
	Create Prediction	·	• •	• •	·	26
0.10		•	• •	• •	•	20

3.11 Set up BigML prediction	27				
3.12 Create Centroid	28				
3.13 Create Anomaly Score	28				
3.14 Create Topic Distribution	29				
3.15 Find Resource	29				
3.16 Select the Custom Value option for the Resource argument	30				
3.17 Select the ID of the resource from the Find a Step action	30				
3.18 Use the Additional data argument to add the resource input fields names and select the					
corresponding data from a previous step	31				
3.19 Map the input fields of the selected resource to the input data from previous steps	31				

Glossary

- Anomaly Detection an unsupervised Machine Learning task which identifies instances in a dataset that do not conform to a regular pattern. 1, 19, 28
- **Classification** a modeling task whose objective field (i.e., the field being predicted) is categorical and predicts classes. 26
- **Clustering** an unsupervised Machine Learning task in which dataset instances are grouped into geometrically related subsets. 1, 19, 27
- **Dashboard** The BigML web-based interface that helps you privately navigate, visualize, and interact with your modeling resources. 22
- **Deepnets** an optimized implementation of deep neural networks, a class of supervised learning algorithms, that can be used to solve regression and classification problems. The input features are fed to one or several groups "nodes", each group of nodes form a "layer". Each node is essentially a function on the input that transforms the input features into another value or collection of values. This process continues layer by layer, until we reach the final output (prediction), an array of perclass probabilities forclassification problems or a single, real value for regression problems. 1, 19, 26
- **Ensembles** a class of Machine Learning algorithms in which multiple independent classifiers or regressors are trained, and the combination of these classifiers is used to predict an objective field. An ensemble of models built on samples of the data can become a powerful predictor by averaging away the errors of each individual model. 1, 19, 26
- **Logistic regression** another technique from the fields of statistics that has been borrowed by Machine Learning to solve classification problems. For each class of the objective field, logistic regression fits a logistic function to the training data. Logistic regression is a linear model, in the sense that it assumes the probability of a given class is a function of a weighted combination of the inputs. 1, 19, 26
- **Model** a single decision tree-like model when we refer to it in particular, and a predictive model when we refer to it in general. 1, 19, 26
- **OptiML** an automated optimization process for model selection and parametrization (or hyperparametrization) to solve classification and regression problems. 21
- **Predicting** the result of obtaining the objective field value for your new data using an existing model. The model returns the predicted value along with a performance measure (confidence for classification or expected error for regression). 1, 21
- Project an abstract resource that helps you group related BigML resources together. 22

- Script a compiled source code written in WhizzML for automating Machine Learning workflows and implementing high-level algorithms. 21
- **Supervised learning** a type of Machine Learning problem in which each instance of the data has a label. The label for each instance is provided in the training data, and a supervised Machine Learning algorithm learns a function or model that will predict the label given all other features in the data. The function can then be applied to data unseen during training to predict the label for unlabeled instances. 19
- **Topic Model** an unsupervised Machine Learning task which identifies the relevant topics in the dataset text fields. Topic models in BigML are an optimized implementation of the Latent Dirichlet Allocation algorithm, a probabilistic method to find topics in large archive of documents. 1, 19, 28
- **Unsupervised learning** a type of Machine Learning problem in which the objective is not to learn a predictor, and thus does not require each instance to be labeled. Typically, unsupervised learning algorithms infer some summarizing structure over the dataset, such as a clustering or a set of association rules. 19

References

[1] The BigML Team. *Classification and Regression with the BigML Dashboard*. Tech. rep. BigML, Inc., May 2016.

