Azure ML

RStudio Web Service API

Demo Guidelines

# Introduction

Currently, millions of Data Scientists and developers are using RStudio to build and train Machine Learning models but they lack the ability to share them via web services. Ideally, a user should be able to publish a web service without ever having to leave RStudio. This “In Progress” project will provide that link between RStudio and AzureML's publishing service in order to encourage RStudio users to utilize AzureML web services. Essentially, this will provide a seamless way of sharing data so that when data does need to be shared, a link can be shared with another user and the data is then duplicated for their use.

The three main goals of this project is to provide the RStudio user with Discovery, Consumption and Publishing abilities. This package will be available to users as a downloadable CRAN as long as you have an authorized workspace in AzureML that is discoverable.

# Publishing

Publishing is a simple one call function for the user. Upon calling, the user simply needs to provide the workspace identification information, the function they want published and the name they want this service to be saved as. The publish function call will handle the API call for the user and any consumption that may need to be done related to this function. Once the call is finished, the function will return a list with the web service details, the endpoint detains and the consumption function to the user.

# Discovery

The discovery code allows the user to retrieve a list of the web services available in their workspace given that they provide the workspace ID and the authorization token (both of which can be found in settings on the AzureML webpage). A user can also get detailed information about a specific web service, retrieve its endpoints, and the details of a specific endpoint.

# Consumption

Consuming a web service currently gives the user the option to score either a CSV file, data frame or individual requests. With either option, the user simply has to make a single-line function call and the scored probabilities are returned to the user in a data frame. There are three functions that the user can choose from: consumeLists, consumeFile, and consumeDataFrame. All functions return a data frame.

# Getting Started

Why are all these Steps necessary? So the user can run the web services without running into problems. There will be a HOW TO document with a walkthrough of the set up. The working directory and workspace information need to be saved locally for security reasons. The reason that the input and output need to be formatted is this is a necessary step for publishing. Before you can call any of the functions, which you will be able to find a guide to all of these in the how to document, you will need to compile all of the code. One of our end goals is to see the CRAN package help the user with this and compile this code for the user and install all necessary packages.

# Summary

Rstudio users can now **Publish, Discover and Consume** trained models from RStudio using AzureML web services across multiple platforms (such as Visual Studio) as long as the user has a valid and discoverable workspace with AzureML.