DATSR User Documentation

Release 1.1

Dux D-zine

Oct 30, 2022

CONTENTS:

GETTING STARTED

WARNING: Due to time constraints the team was unable to implement certain features of this application. The functionality exists in both the front and back end, but interfacing and web hosting brought unexpected hiccups. See the walkthrough section (*Walkthrough*) for more information on missing features. We apologize for this inconvenience.

Welcome to DATSR! This document is to help you get a little more acquainted with our website and guide you through the use of our revolutionary online time series repository.

Note: For questions about this document or anything else related to DATSR, feel free to get in touch with the team at duxdzine@not_real_mail.com or by calling our help line at 1-800-867-5309.

Now let's go through each of the sections in the user documentation to get a better idea of the content we'll be covering.

1.1 "Overview" Section

In this section we give an overview of the functionality of DATSR and the main ways to use it.

1.2 "The Site" Section

This section of this document describes the web interface which acts as the frontend to the DATSR application. Here you will find descriptions of the individual pages, tutorials on how to interact with the user interface, and general site navigation.

1.3 "Input Formatting" Section

In "Input Formatting" we describe how inputs into the DATSR system should be formatted. This includes instructions for formatting files containing predictions or new data sets as well as formatting for filling out forms available on the site.

1.4 "Scoring" Section

Here you will find a more in depth guide to how the scoring works on DATSR. You will find the statistical methods that were used for the ratings generated as well as justification and a little bit of background for the mathematics behind our calculation.

TWO

OVERVIEW OF THE SYSTEM

DATSR is first and foremost an online repository built with time series data in mind. This means our site goes beyond simple data storage by providing features specific to TS datasets such as hierarchical/set structures, automatic validation/testing set division, and metadata specific to TS data sets.

In addition to providing a repository for time series data, DATSR also has a built in system for testing predictions made on time series. The site allows you to upload their data to see the effectiveness of their predictions not just in the score reflected, but also by ranking you among everyone else who has submitted predictions for that specific data set.

THE SITE

The website for DATSR is composed of four pages each providing unique functionality and information to you the user. The navigation bar at the top can be used to seamlessly switch between the different parts in the modularized application.

The details of each page's functionality and interface are given below:

3.1 Home Page

This page can be reached by clicking the app name in the top left corner of the site. On it, we provide an overview of our application including its functionality, the names of our team members, and our story. Also available on this page is a link to the html version of this user documentation for convenient access.

The home page is where users will land when they first visit the site and is meant to orient them to the application's interface as well as provide easy access to other pages through the navigation bar.

3.2 Score Board

This page displays the "scores" that users of this application have achieved by entering their predictions using the "Enter Your Predictions" section of our site (See [LINK]).

To use this page simply select the time series data set you would like to view the high scores for and the page will refresh to show you the top predictions and their scores.

At the top you can see the 5 users with the highest score for that particular dataset–their success has been rewarded by placing boxes around their names. Below that there is a section for other users' scores who were not able to make the top 5.

3.3 Upload Your Files

This page is where users can upload either new datasets to be added into the repository or predictions they have made for datasets they have previously downloaded and built predictive models for.

To upload a new dataset into the repository users should. . .

To upload predictions for an existing time series data set users can. . .

3.4 Databases

On this page you can find all of the time series data sets available in the repository. For each data set, meta data is displayed that helps you to narrow down which data set you would like to download. Each data set also has a download button which automatically adds a .csv file containing the training data for the set to your browser's downloads folder.

WALKTHROUGH

This section is intended to give a short tutorial that will familiarize new users to DATSR's interface and functionality. First, to do this demo please have the following ready:

- The file EARTHQUAKE_OCCURENCES_2015.csv which was included in the zipped folder that this project was submitted in

- The file chua_circuit.csv which was also included in the zip file

- Access to a web browser to reach the DATSR site

4.1 Step 1: Navigating the Page

Once you enter the URL for DATSR into your browser, you will be brought to the home page where you can read more about the project, the team, and access the user documentation (this document) if you please. Notice the search bar at the top of your browser and click through the 4 links (the DATSR logo is a working link as well). This gives you an idea of how to get around the broad structure of the DATSR site. For more information on each of the pages that you can visit, see the "The Site" section of this user documentation (Section 3).

4.2 Step 2: Downloading a Dataset

Go to the "Databases" page using the navigation bar at the top and browse through time series data sets available in the repository. When you have found one that piques your interest, click the download button to add the .csv file with the set's training data to your browser's downloads folder.

Note: The remaining two steps of this walkthrough demonstrate **intended functionality** that we were unable to implement due to time constraints.

4.3 Step 3: Entering Predictions

For the sake of this demo, it is best to simply generate random data for predictions unless you already have a predictive scheme set up to efficiently make predictions for the validation set of the data you downloaded. Either way, predictions should be written in a .csv file and formatted in a way that complies with DATSR's backend. For more information on the proper formatting conventions, see the "Input Formatting" section (Section 5).

After you have your "predictions" for the validation part of the set completed, navigate to the "Upload Your Files" page on the site. Here you will be filling out the form that is on the top of the page. Enter your name (mandatory), the GitHub

URL for the predictive scheme used (optional), the collection that the data set came from, the name of the dataset, and the .csv file using the "Choose File" button.

After you have filled out the form, the site will display your "score" which is equal to the calculated Mean Square Error (MSE) of your predicted values compared against the actual values in the validation set. For more information on scoring see the "Scoring" section of the user documentation (Section 6).

Now navigate to the "Score Board" page and select the data set that you made a prediction for. The page will then show the score board for that particular data set and you can see how your prediction matches up with your peers.

4.4 Step 4: Uploading a New Dataset

DATSR also allows for users to contribute to the repository by adding their own datasets to be made available to other users. To demonstrate this functionality, it will be handy to have easy access to the two .csv files mentioned above (EARTHQUAKE_OCCURENCES_2015.csv and chua_circuit.csv).

Once again, navigate to the "Upload Your Files" page, but this time you will be filling out the lower of the two forms ("Upload a Dataset"). In the "Dataset" field enter a name for the data you will be uploading. Then, using the "Choose File" button again, select the .csv file you wish to upload into the repository. When you click "Submit" the data will be sent to the backend database to be made available other users.

FIVE

INPUT FORMATTING

Data inputted into DATSR—whether it is data for a new dataset or predictions submitted for an existing data set—should be in the form of a .csv file. The left-most column(s) should be populated by time values and to their right there should be the values for whatever variable is being measured over the time period. If there are no time values for the time series, the first column should be the values that are being measured.

This formatting standard was selected because it is a common way for time series to be stored in other repositories. When we built DATSR, we wanted it to fit in well with the established research community and this formatting standard adds to DATSR's compatibility with the systems already in place.

Data downloaded from the repository will be formatted in this way, but other examples of correctly formatted data can be found in the zipped folder that this project was submitted in (chua_circuit.csv and EARTHQUAK_OCCURENCES_IN_2015.csv). For more information on these particular data sets, see the demo section of this document (Section 4).

SCORING

DATSR rates the predictions submitted to the site using Mean Square Error (MSE)—a common error calculation in classical statistics. The following equation shows how Mean Square Error is calculated on a set of n actual values x_i and predicted values y_i:

$$MSE = \frac{1}{n}(x_i - y_i)^2$$

Fig. 6.1: Mean Squared Error Calculation

MSE was selected as the statistic used to rank the predictive models in DATSR because it is common in industries that use time series analysis, as well as in the research community.