**Executing Source Code of AMR Association Mining Project**

This document describes the steps of executing AMR association rule mining source code. The code belongs to Cazer Lab at Cornell University. For technical details contact Abdolreza Mosaddegh (ORCID: 0000-0001-5840-3628) – Email: am2685@cornell.edu

Association Mining of Microbe Resistance source code consists of the following Python classes:

*Common\_functions.py, data\_base.py, dtree.py, rule\_graph.py, rule\_set.py, transaction\_set.py, and spark\_fp\_growth.py*

The code is available from <https://github.coecis.cornell.edu/am2685/HorseData>

**Prerequisites:**

Runtime environment of Python 3.4 or later is required.

SPARK 3.3 or later should be installed on the client.

MS SQL Server 15 (or later) is required.

All these platforms should be installed as described in **Setup\_Bigdata\_Platform\_Manual** document**.**

**Step 1: Restore Target Database**

Open SQL Server Management Studio. Right click on *Database* and select *Restore Database*.

A screenshot of a computer

Description automatically generated

Then add the backup file of HorseDB database located in *Database* directory in *Device* section*.*

A screenshot of a computer

Description automatically generated

**Step 2: Install Python Libraries**

The following python libraries should be installed:

* pyspark
* pyodbc
* regex
* scikit-learn
* graphviz
* os-sys
* networkx
* matplotlib
* statistics
* holoviews
* bokeh
* scipy

For installing the libraries open a new command-prompt window using the right-click and **Run as administrator**.

If PIP is not installed, use the following command for installing PIP.

py -m ensurepip --default-pip

Then use PIP for installing the libraries using the following command:

py -m pip install [Library Name]

**Step 3: Run Python Code**

The main class for the starting of process is *spark\_fp\_growth.py* located at *Codes\PythonCode* folder. Configuration parameters are also defined in this class.

For setting the parameters, open *spark\_fp\_growth.py* using Notepad or any other text editor and change the parameters according to the Note section in code\_output\_checklist.xlsx file then save the changes to the file.

For executing the code, open a new command-prompt window using the right-click and **Run as administrator** and use the following command:

py [path of the source code]\spark\_fp\_growth.py

By running the code, extracted association rules will be recorded in the database and charts will be rendered and displayed to the user. After displaying each chart, the process waits for the user to close the chart and then moves to the next step.

**Step 4: Retrieve Results from the Database**

To retrieve results from the database, open SQL Server Management Studio and open each SQL script located at *Codes\SQL\_Code* folder.

Settings and outputs of each SQL script and corresponding paragraph, table, or figure in the paper is described in code\_output\_checklist.xlsx file.