Test Plan:
Repository and Mining of Temporal Data

Team Members:

Jessica Nguy jnguy2014@my.fit.edu
Siomara Nieves snieves2014@my.fit.edu

Faculty Sponsor and Client:

Philip Chan pkc@fit.edu

Document Outline

1. Introduction

This document describes the test cases for Repository and Mining of Temporal Data (RepoMining). The document details how to verify that the system is operating correctly. This document provides a brief summary of necessary items needed for each requirement before listing test cases. For more information on each requirement, refer to the requirements document.

2. Definitions

GUI - Graphical User Interface. It is what the user sees when using the program.

CSV - Comma Separated Value. It is the data format accepted by the program.

Repository - A central location where data is managed.

Data Providers - Users who upload data to the site via a .CSV file. Values in data possess timestamps either in vertical or horizontal format, and the data must be numerical.

Data Consumers - Users who can use the data on the site to answer questions. A data consumer can be a data provider also. Data customers must specify a target variable.

Q1 - Shorthand for “Is there a significant change in the target variable?”

Q2 - Shorthand for “Why was there a significant change, and what are the top variables that are correlated with the target variable?”

Q3 - Shorthand for “What is the value of the target variable in the next timestamp?”

Meta-data - Descriptor that contains the description of each variable, start/end timestamps, tags, and a public/private setting.

Correct Format - csv file that has timestamps going across the first row and variables listed in the first column, or timestamps going down in a column and variables listed in the first row.

3. Test Plan

3.1. uploadData

- Software is able to accept and read .csv files
- Checks whether or not the csv file is in the correct format
- Able to output a csv file with the results from answering Q1, Q2, Q3
- Must be able to store a short description of data, who uploaded/authored the data, and what type of access it has.
- Data is stored in the program.
- To verify that uploadData works, read the number of occurrences of a variable, print out an error message if csv file is in the incorrect format, display meta-data when prompted.

3.2. questionOne

- Program is able to answer Q1: Is there a significant change in the target variable?
- Able to calculate the average rating for variables that correlate to the target variable, average change in rating, standard deviation, z-score, and compare the result to answer Q1.
- To verify that Q1 was answered, calculations must be correct. Output must match with a known example that is correct.

3.3. questionTwo

- Program is able to answer Q2: Why was there a significant change, and what are the top variables that have affected the target variable?
- Able to calculate and use Pearson Correlation and Cross Correlation Program to determine the variables affecting the target variable and significant changes in the target variable.
- To verify that Q2 was answered, calculations must be correct. Output must match with a known example that is correct.

3.4. questionThree

- Program is able to answer Q3: What is the value in the next timestamp?
- By using Linear Regression and Cross Correlation, the program will be able to forecast the next value in a defined set of time.
- Display results in a graphical form.
- To verify that Q3 was answered, calculations must be correct. Output must match with a known example that is correct.

3.5. setTag

- Program is able to set user-defined tags as additional information on the uploaded document.
- To verify that setTag works, print out the tags associated with the document.

3.6. setTarget

- Program will set a user-defined target variable as the target variable to analyze.
- To verify that setTarget works, use variable in a sentence. Also be able to use variable to search through repository.

3.7. narrowData

- Software will narrow down the search of correlated variables.

- Use Pearson Correlation and Cross Correlation to determine if target variable is highly correlated with values.
- To verify that narrowData works, make sure calculations are correct. Use a correct example to check program output.

3.8. visualizeData

- Software will take in results from questionOne, questionTwo, and questionThree to display the results in graphical form.
- Software will prompt user to select what variables user would like to download/keep in checkbox form and allow for a csv file download.
- To verify that visualizeData works, graph must be correct by inputting a known example and comparing results. When variables are selected, software will allow users to download file that corresponds to the chosen variables.

3.9. metadata

- Software will prompt user to add a short description to the data, start and ending timestamps of the csv file, tags, and whether or not the data is for public or private use.
- metadata will store meta-data in database.
- To verify that metadata works, user entries must be saved to data. Entries will be printed out to ensure that entries are saved correctly. Retrieval and storage of meta-data in database will be checked.

3.10. webApp

- Program is hosted on a web application
- Web application must have a target variable bar, a csv upload button, a search button.
- Application will have 6 different pages, one for Data Providers, Data Consumers, Answer Q1, Answer Q2, Answer Q3, and display the results graph at the end of the results.
- To verify that webApp works, all buttons and displays must work. Each button will go to the next page. Error messages will display if there is no target variable input.

3.11. exportCSV

- Program will export .CSV file with data about the target variable, the top-k variables and their correlation to the target, and the value generated as the next possible timestamp.