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

Meetings With the Client:
August 25; September 5

Goals:
Our project aims to simplify the appearance of data mining by developing a generic, basic tool that is able to analyze and predict future trends. We want to create a program that allows users to access data in one place without having to visit multiple websites to get answers to their questions. The program will allow data providers to share their data amongst other data providers and users on the website application. They will also be able to as use the program to analyze correlations between other variables and the target variable. The program will also be able to recognize that some data sets have different timestamps and can start at different time periods, so the program must have the ability to integrate data that has been recorded in different time intervals. Speed is an important factor to analyzing data and will be a focus for our project. The entire project will be on a web application.

Approach:
The project is meant to answer three questions.

Is there a significant change in the target variable? As a repository, the program will analyze data collected from data providers, allowing the user to see whether or not there was a change in the target variable.

Why was there a significant change, and what are the top variables that have affected the target? An important feature of a good analyzer is the ability to list the variables that affected the target query the most. This allows for the program to narrow down its search and further refine the changes made and analyze it; in addition, the program will show the factors that have impacted our target during a certain period of time. With the large amount of variables that will be analyzed, speed and optimization is one of our main areas of focus.

What is the value in the next timestamp? Not only will the program be able to analyze the data, but it will be able to predict what will happen within a given time period, such as predicting the next year. All three of the user's questions should be answered within the application. We plan to have an option to export a

.CSV file once the analysis is done. The end output of the .CSV file will allow for another program to access what our program has analyzed and predicted.

Novel Features:

A novel feature in our project is the simplified user interface that possesses a few buttons. The program will also have an analysis component to detect and predict the values that influence the target variable. The application will be able to recognize and fix problems with data with different timestamps, or timestamps that have multiple points of data.

Technical Challenges:

The team members are uninformed on what exactly data mining entails along with what algorithms to use to analyze data. We want to be able to have all the data analyzed quickly, so speed and optimization is one of our main focus. Optimizing speed is vague; further research is required. Although reducing the number of variables to increase the program speed can be beneficial, deciding on which variables to skip over would be challenging. We want to focus our efforts into finding an algorithm to analyze data, and an algorithm to store and retrieve data.

We also do not know which language to use for the application, after all, we want to use a language that works quickly at analyzing data, but is easy to pick up. Currently the languages we are looking at are Python, Java, C++, and R. Languages for the website application itself is a debate between HTML, JavaScript, or CSS. We are open to using multiple languages in the same program if the main language we choose lacks a certain feature.

One of the goals of the project is to make a simple interface for the user to use. To do this we need to research how to create a GUI to 'hide' all the analyzing and code going on in the program. Several inputs are planned: a text box to insert the query, a 'Start' button, an exit button, a button to download or open the resulting .CSV file, and a button to upload the specific .CSV file should the user decide to upload one. The minimize and maximize buttons are optional for the current plan. Along with the GUI the team members do not know how to make a website and whether or not additional plugins are required. Research into website creation is needed, as well as sites to host the application.

Milestones:

Milestone 1 (Oct 2): itemized tasks:

- Investigate which programming language is best for our project, and learn how to use it if needed
- Determine if additional plugins are needed, research algorithms, web application plugins
- Preliminary program able to read .CSV files
- Create Requirements Document
- Create Design Document
- Create Test Plan

Milestone 2 (Oct 30): itemized tasks:

- Design, implement, test, and demo GUI for user input (demo can display query back)
- Implement, test, and demo whether or not the target variable has a significant change (Question 1)
- Get website running
- Implement, test, and demo .CSV file reading
- Implement, test, and demo preliminary analysis

Milestone 3 (Nov 27): itemized tasks:

- Refine analysis algorithm
- Refine website
- Implement, test, and demo which variables correlate with the target variable the most (Question 2)
- Add additional buttons on GUI if needed
- Implement .CSV output

Task Matrix for Milestone 1:

Task	**Jessica**	**Siomara**
Investigate/Select Tools	Languages, HTML	Languages, CSS, JavaScript
Investigate additional plugins/sites	Plugins, Research Algorithms, How to make Web Applications	Plugins, Research Algorithms, How to make Web Applications
Create .CSV program	Create test .CSV cases	Create program
Requirement Document	75%	25%
Design Document	50%	50%
Test Plan	60%	40%

Approval from Faculty Sponsor

"I have discussed with the team and approve this project plan. I will evaluate the progress and assign a grade for each of the three milestones."

Signature: _____ Date: _____