UNIVERSITY OF SOUTH FLORIDA

COMPUTER SCIENCE AND ENGINEERING

Name: LALITHABHINAYA MALLU

ID: U63881688

**OBJECTIVE**

To find out degree of distinguishability across weeks between different users and single user.

**TECHNICAL DETAILS**

I have developed project using Visual Studio2015 with .NET Framework 4.6.1 in C# language. This project is designed as a console application. Below Nuget packages (references) are included to develop this project.

**1. Excel Data Reader:** Reads complete excel file data into data table.

**2. Excel Data Reader. Dataset:** It converts data table into dataset(objects).

**Procedure:**

1. This program takes folder path as input and reads data from all 54 excel files.
2. Data is stored in table object and then time intervals can be divided based on bins.

* Used below formula to divide the time interval.

Bin(i) = floor(k\*(x(i)-min X) / (maxX-minX))

* Here I observed that, we don’t need to calculate real first packet epoch values into human readable format. We can directly calculate time intervals based on epoch values.

1. I have observed that while calculating spearman’s coefficient for same user, r1a1a and r2a2b produces value as 1 and Z-value results produces indefinite value. Here professor instructed me to modify R1a2a and R2a2b as 0.999.
2. Based on spearman’s coefficient values I have calculated Z- value.
3. Finally produced P value from the given method and result calculated as 1-P.

**Steps to run program**

1. Download the ZIP file and extract the contents in the folder.
2. Try running executable which is located in (ISAssignment\bin\Release\ISAssignment.exe).
3. When you start executing the program, console prompts you to provide folder path.(folder which has 54 excel sheets)
4. Once you provide input path, press Enter.
5. Then application starts reading files from the folder and console prints P value for all the excel files.
6. You may change time intervals to test the code for remaining time intervals.

**OBSERVATION**

In this project, I took three time windows (5 minutes, 227 seconds and 10 seconds) to profile user activities for two weeks (week 1 & week 2). I observed that when time interval is 5 minutes, degree of distinguishability across weeks between different users is very less. The below sample space shows multiple gray cells which have P value greater than 0.05. This depicts that in large window different users have lot of similarities while browsing internet. Also, as time goes on similar user has noticeable degree of similarities (diagonal entries-printed in blue color).

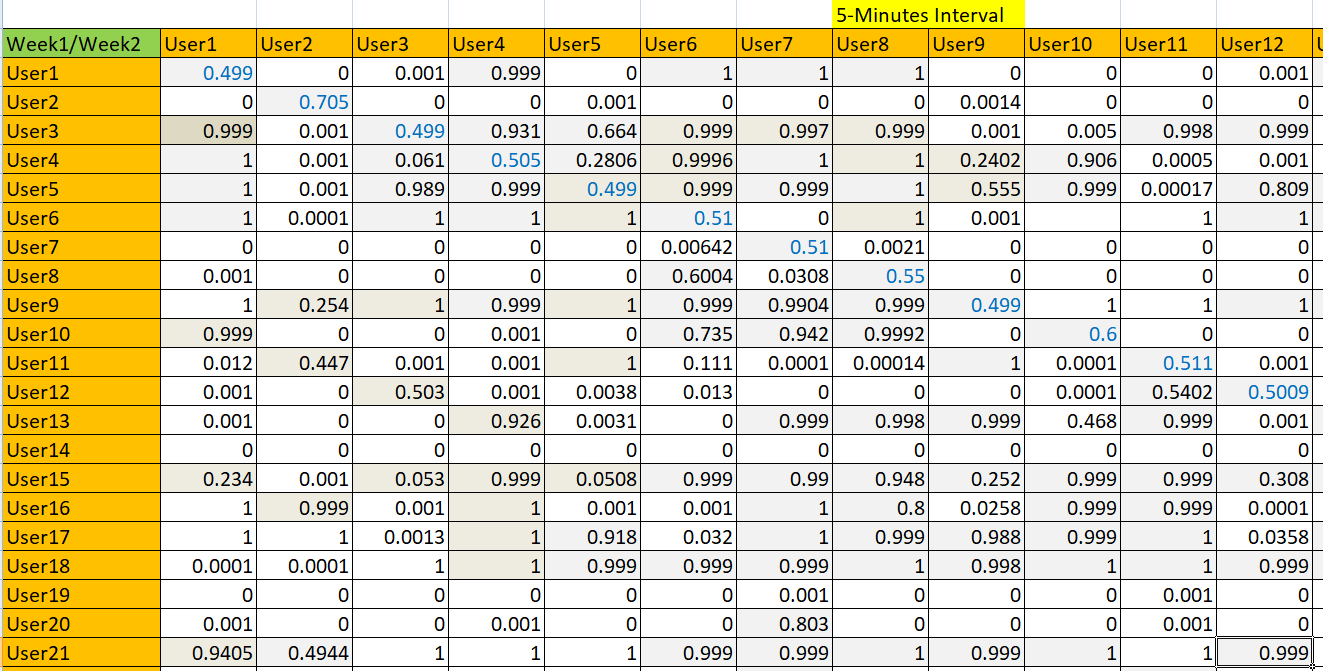


Figure 1: Sample users report for five minutes interval

I would like to like to compare gray cells in other time intervals (227 seconds and 10 seconds). In 10 seconds interval, different users have P value less than 0.05. Which depicts that degree of distinguishability across weeks between different users is more and also same users across weeks have dissimilarities. I personally feel that profiling users with very less time window (10sec) is not preferable.

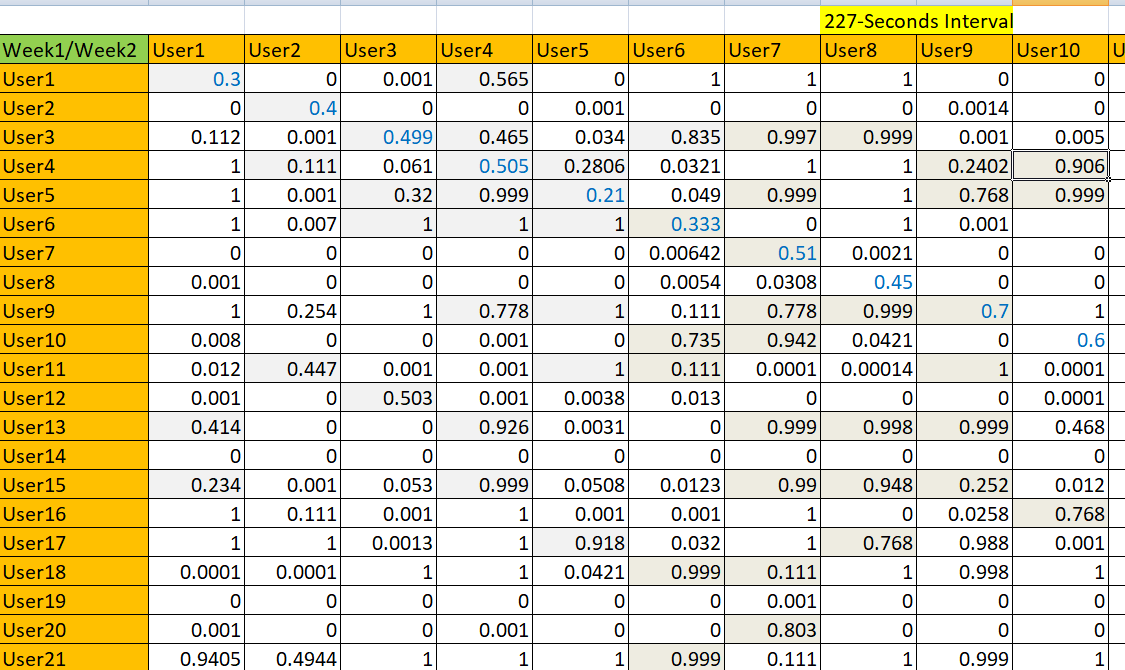


Figure 2: Sample users report for 225-Seconds interval

Also as per my observation, when compared to 5 minutes and 10 seconds time intervals I would suggest to take 227 seconds which has reasonably good amount of similarities between same users across weeks. There is dissimilarity between different users across weeks in 227 seconds window.

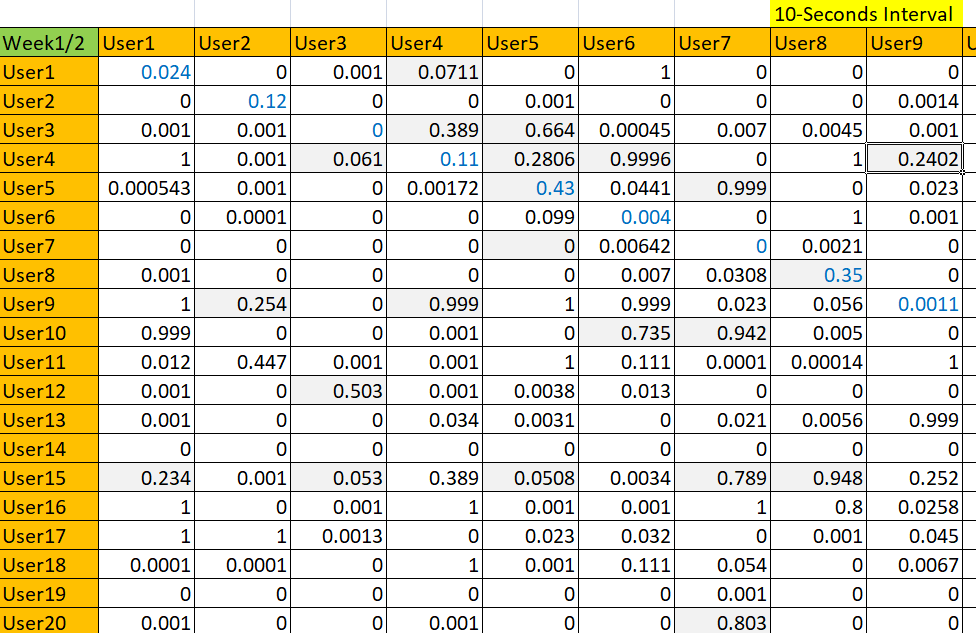


Figure 3: Sample users report for 10-Seconds interval

**RESULTS**

Results of each time window is stored in three different spread sheets. Find below reports for complete P-values.

5-min Interval 225\_seconds interval 10\_Seconds interval