RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Personal Health Companion

By:
AKSHITA AMBATI
HARIKA MATTA
JIANYU ZHANG
RUIQI LIN
SHIKHA KAKAR
YUEYANG CHEN

INTRODUCTION

- Lack of awareness of:
  - the need for a healthy lifestyle
  - The adverse impacts of unhealthy habits
  - Work out styles and diet

- Lack of motivation among people to continue following a healthy lifestyle
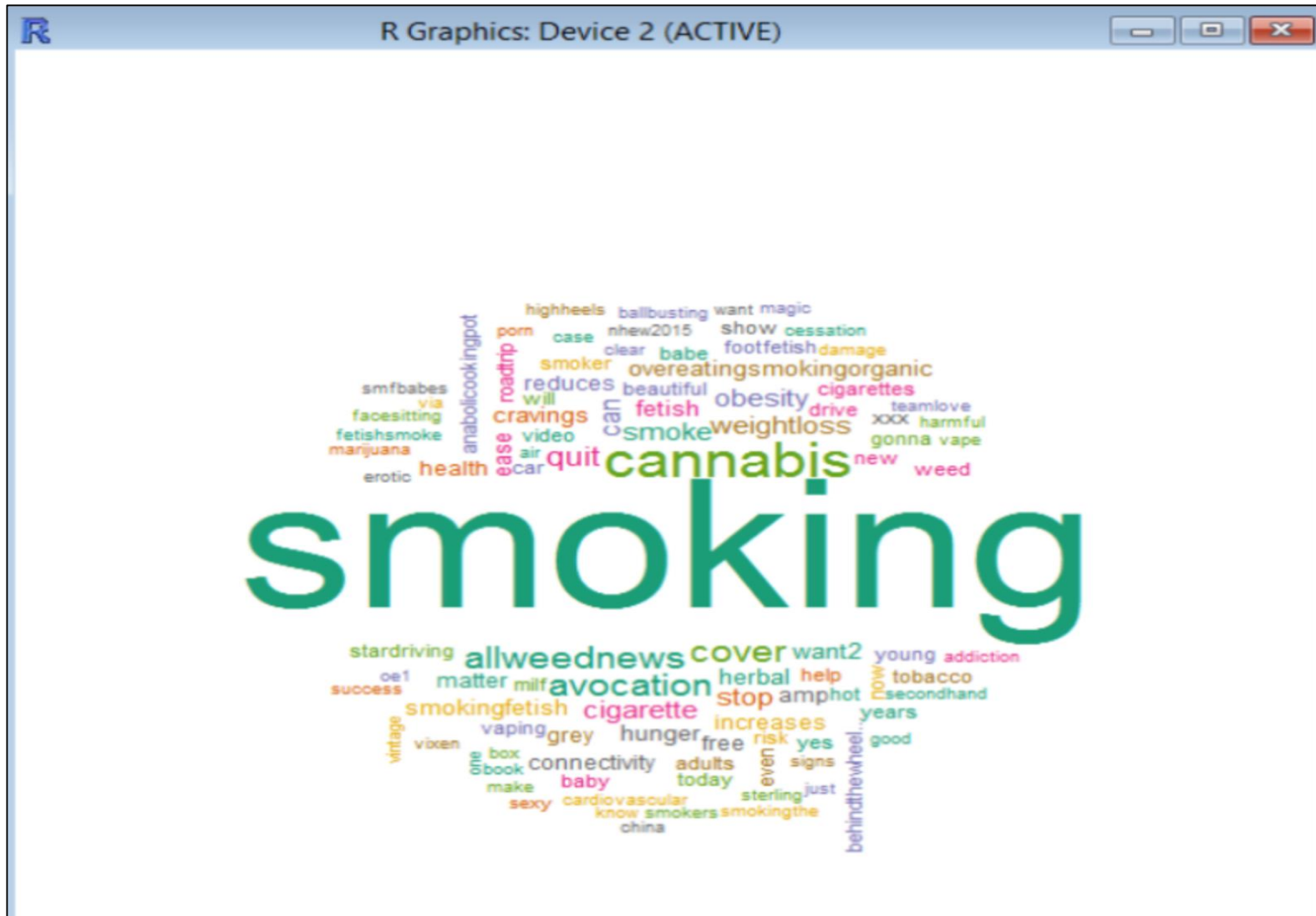
Features

- Language Analysis (using LingPipe) of Individual tweets for:
  - Weekly Diet Trends
  - Weekly Exercise Trends
  - Weekly Alcohol and Smoking Trends
  - Colder vs. Warmer – weekly trends
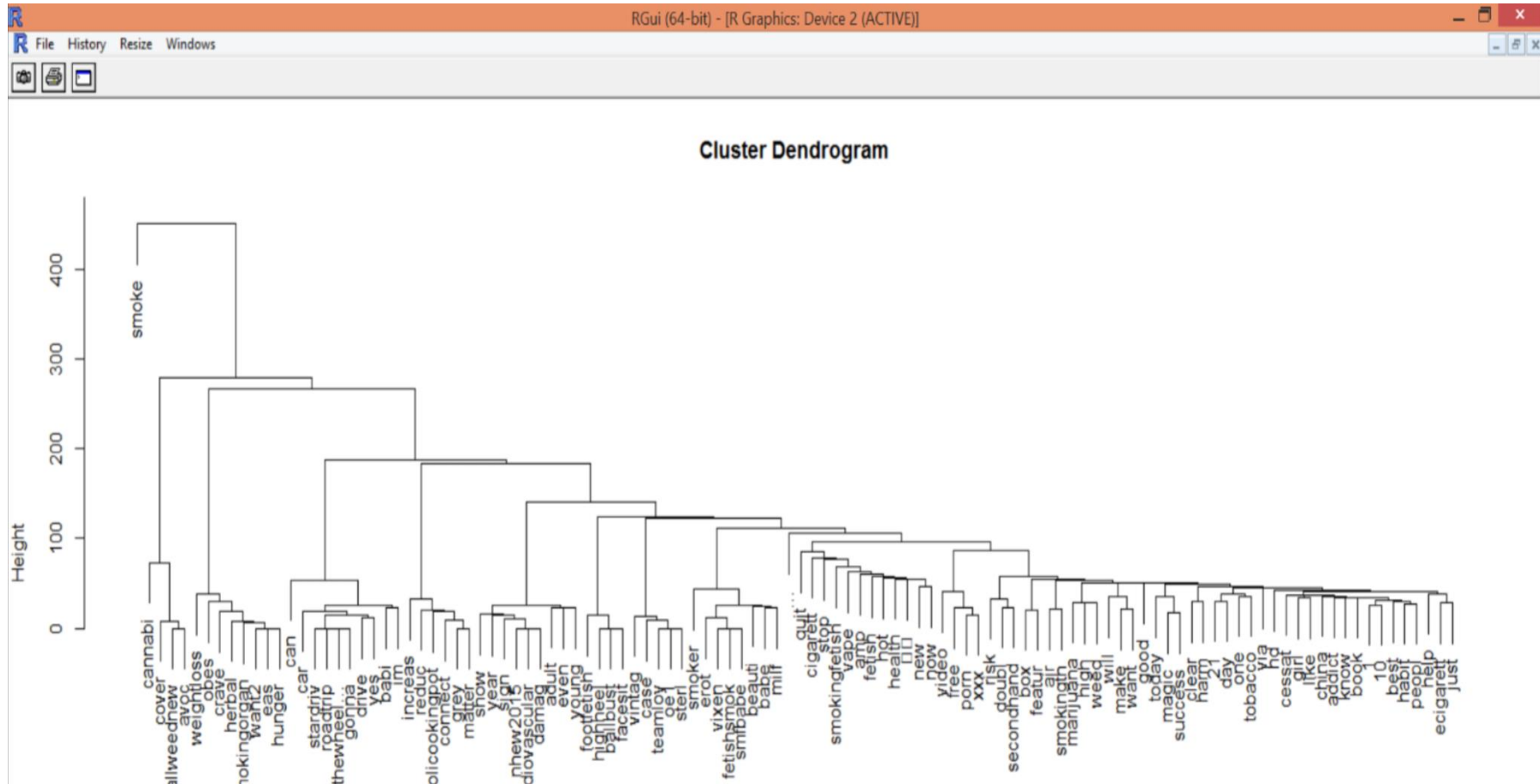
- R – for clustering of relevant tweets
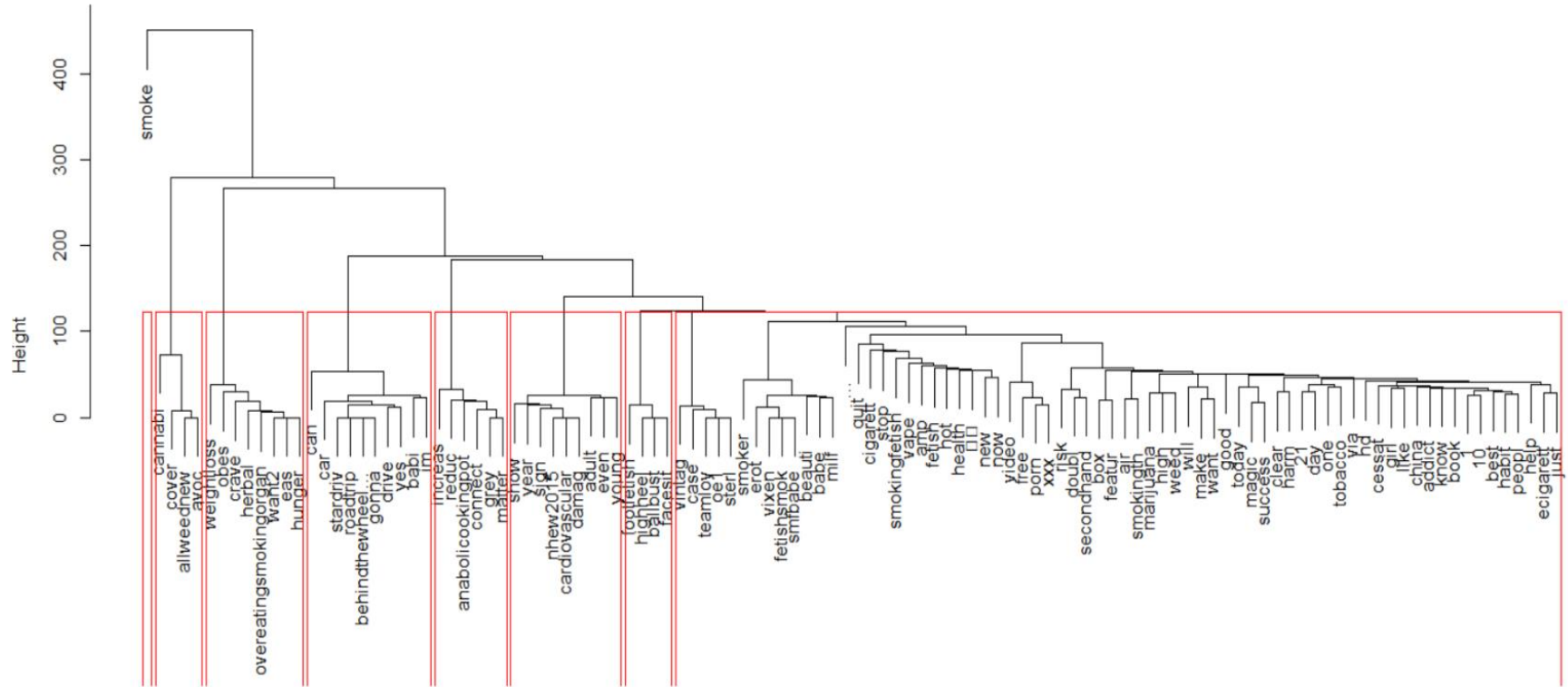
Dietary Word Cloud
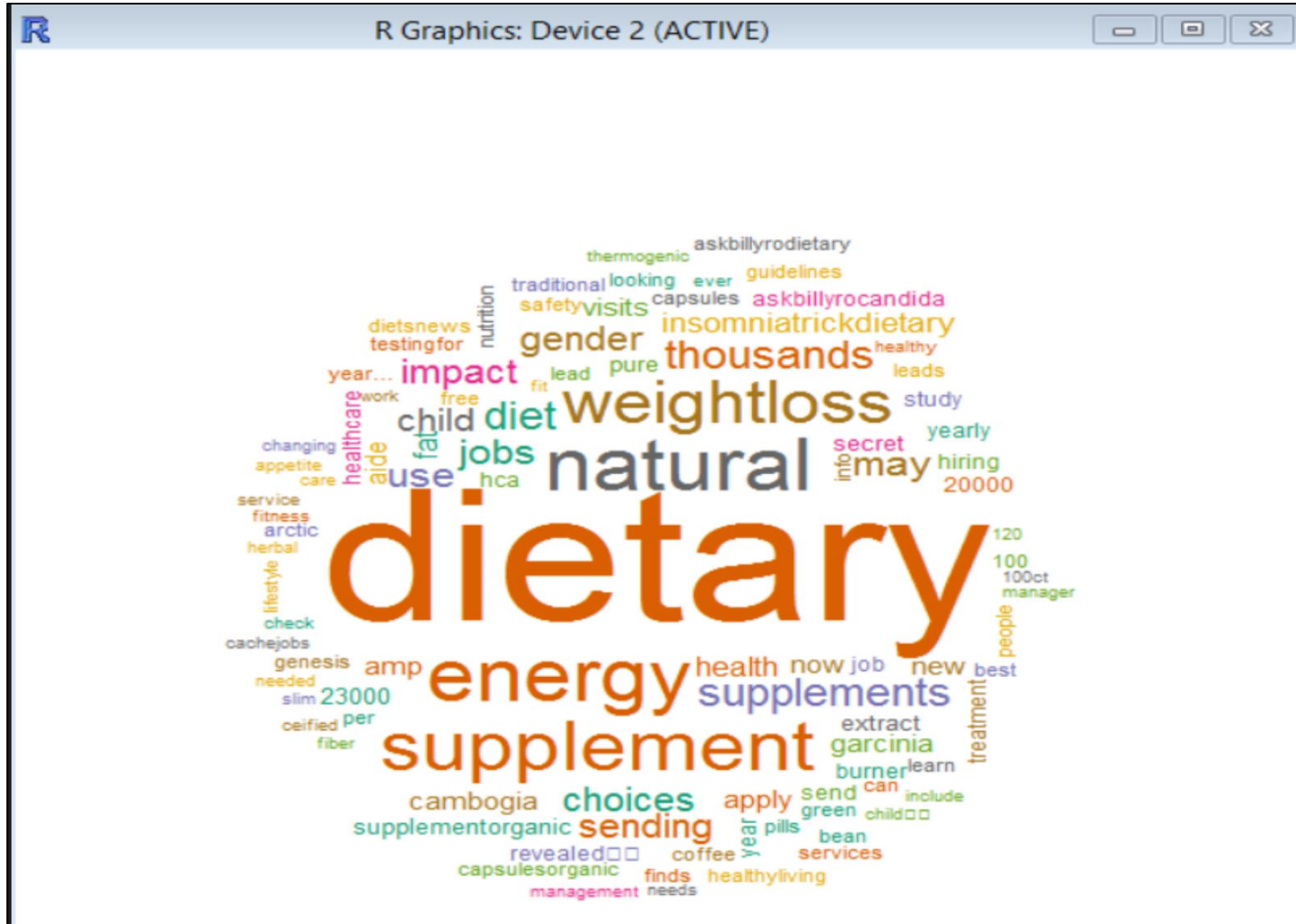
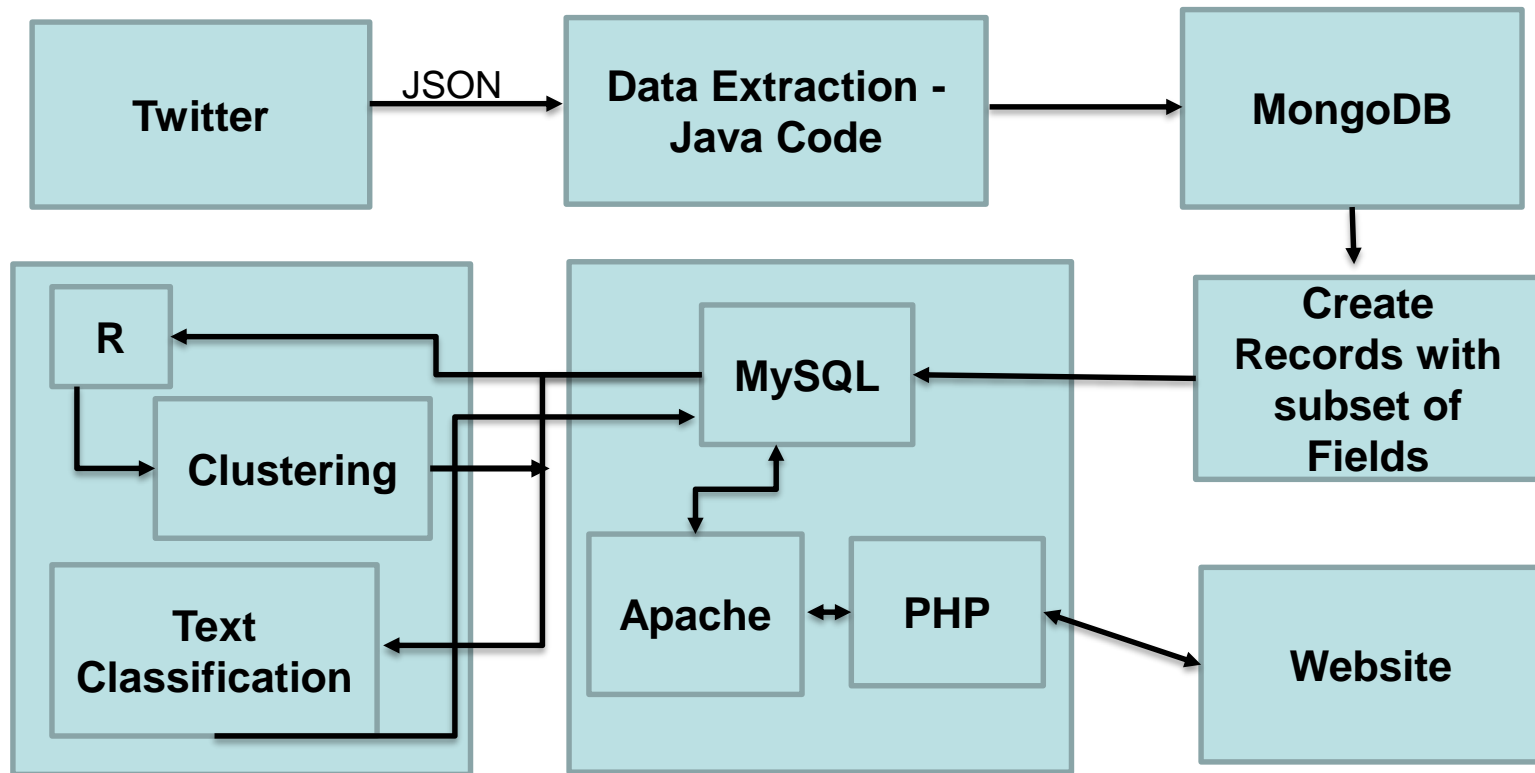# Smoking Word Cloud

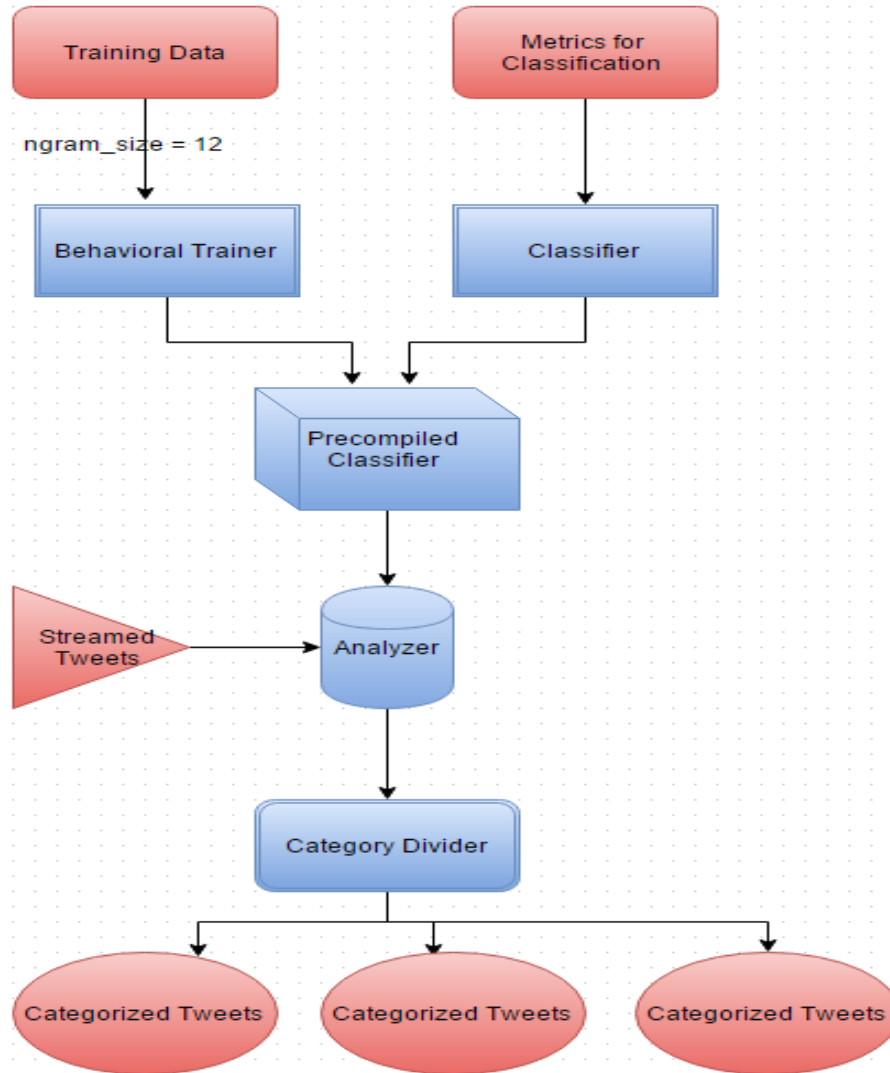# Clustering Dendrogram

Clustering Lines

Cluster Dendrogram

Data Analysis

Data Analysis

Categories:
- pos.exercise
- mot.exercise
- pos.diet
- mot.diet
- neg.diet
- pos.smoking
- quit.smoking
- pos.alcohol
- quit.alcohol
- garbage.all

Testing

- Function Tests
  - Tweet Collection
  - Histogram Display

- Integration Test
  - Unit Tests run on daily basis – as soon as the code is written.
  - Code – committed only if it passes the unit tests.
  - Acceptance Test - every 1-2 weeks.

Backup Slides

Garbage Tweets

527566554267807744	Entrenando con demasiado calor #GYM espalda y tric...	2756704688
527566558722146304	#WeightLoss #Program Health Tip: Practice a Well-Balanced Exercise Program http://t.co/UfsPpeVgGM #Diet	534928178
527566559783313409	#Healthy #Nutrition How to Lose Weight With Thermo...	598370918
527566559921729536		59369958
527566561129672704		01794252
527566562564132864		69756746
527566563617284096		73865263
527566565911179264		34182600
527566584429428736		39855189
527566588455559168		8040902
527566595267510274	Hoy recorrimos 10 kilómetros de #running con temas...	384937373

#WeightLoss #Program Health Tip: Practice a Well-Balanced Exercise Program http://t.co/UfsPpeVgGM #Diet

Press escape to cancel editing.

Running Tweet

Quit Smoking

Database Table

	TweetID	TweetText	UserID	CreateTime
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566517672890368	Pilateando en Olimpic !! Ya no puede entrar nadie ...	2358627272	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566523301642240	Pilateando en Olimpic !! Ya no puede entrar nadie ...	2358627272	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566533946793984	Buenas noches familia! #SoyDeAgua #shark #surf #wa...	2358627272	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566538640203776	If you want to lose weight contact http://t.co/oRh...	2365562754	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566540531838977	Because what's a day with out my weight loss coffe...	88336290	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566554267807744	Entrenando con demasiado calor #GYM espalda y tric...	2756704688	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566558722146304	#WeightLoss #Program Health Tip: Practice a Well-B...	534928178	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566559783313409	#Healthy #Nutrition How to Lose Weight With Thermo...	598370918	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566559921729536	#EmmaWatson #Fitness #Diet ###Fitness Fitness on t...	559369958	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566561129672704	#FatLoss #Healthy 5 Secrets To Weight Loss While G...	601794252	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566562564132864	#Workout #GetFit Tips to Lose Fat http://t.co/erCn...	769756746	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566563617284096	Después de una sesión de gym toca una duchita fres...	573865263	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566565911179264	#JohnMayer #Fitness Paleo Diet Success Stories htt...	784182600	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566584429428736	8 Min Abs The Classic - Level 1 #workout #abs #six...	189855189	2015-10-30 12:43:24
☐ ✏ Edit ⚃ Copy ⊖ Delete	527566588455559168	Bem melhor assim... #Run #Mar #Floripa #VemVerão #...	98040902	2015-10-30 12:43:24

Sample Tweet Classification

```
----------------
Testing on Tweet : I ran 3.17 mi with @MapMyRide.  #run #running http://www.mapmyride.com/workout/811185395
Got best category of: pos.exercise
Rank  Category  Score  P(Category|Input)   log2 P(Category,Input)
0=pos.exercise -2.8527878883654347 1.0 -262.45648572962
1=garbage.all -4.134556830482898 3.1748066669032881E-36 -380.3792284044266
2=mot.exercise -5.472824622120887 2.746075747642947E-73 -503.4998652351216
3=pos.alcohol -5.8499575541069015 9.86521656797287E-84 -538.196094977835
4=pos.diet -6.204123501972317 1.5330836026502418E-93 -570.7793621814532
5=quit.smoking -7.22456193441882 8.409472685313616E-122 -664.6596979665314
6=quit.alcohol -7.360285168829972 1.46536702687128897E-125 -677.1462355323574
7=neg.diet -7.787121250333351 2.212196743493594E-137 -716.4151550306683
8=mot.diet -7.810135402614176 5.0986163443982124E-138 -718.5324570405041
9=pos.smoking -8.31873640164336 4.186681591308139E-152 -765.323748951189
```

pos.exercise

Sample Tweet Classification

```
--------------
Testing on Tweet : Seriously need to start the #gym
Got best category of: mot.exercise
Rank  Category  Score  P(Category|Input)  log2 P(Category,Input)
0=mot.exercise -1.496112393251283 0.9999991238499288 -50.86782137054362
1=mot.diet -2.100313123366865 6.5462585515056686E-7 -71.41064619447341
2=garbage.all -2.1462898242995343 2.2152395933099748E-7 -72.97385402618417
3=pos.exercise -2.7256835237995154 2.602044311942475E-13 -92.67323980918353
4=neg.diet -3.2069670654907974 3.0858021994177213E-18 -109.0368802266871
5=quit.smoking -3.330557216229261 1.6765651632222177E-19 -113.23894535179487
6=pos.alcohol -3.53548912975657 1.3394861716678278E-21 -120.20663041172338
7=quit.alcohol -3.565592969801908 6.589080811676464E-22 -121.23016097326487
8=pos.diet -3.7619455100165937 6.443957786874213E-24 -127.90614734056419
9=pos.smoking -3.7664197399929567 5.7990767170013B5E-24 -128.05827115976052

--------------
```

mot.exercise

Data Analysis

- Training tweets are used to train the analysis tool
- Tweets are then stored in separate categories using the results of the analysis tool
- Database tables in SQL will store the number of tweets for each category
- Through PHP, the graphs are displayed in the website