# Milestone 1 Progress Evaluation

## Student Performance Analysis

Shreyas Ugemuge sugemuge2014@my.fit.edu

Yaqeen AlKathiri yalkathiri2013@my.fit.edu

Mohammed AlHabsi malhabsi2013@my.fit.edu

Shiru Hou shou2015@my.fit.edu

Faculty Sponsor: Dr. Phillip Chan pkc@cs.fit.edu

March 27, 2017

# Contents


**1 Progress of current Milestone** **3**

1.1 Discussion of each task . . . . . 4

1.1.1 Finalize Conceptualizing 8 behaviors . . . . . 4

1.1.2 Finalize packages, languages and tools . . . . . 4

1.1.3 Implement and test framework to get input from syllabus . . . . . 5

1.1.4 Select data mining techniques for each behavior . . . . . 5

1.1.5 Clean and prepare data . . . . . 5

1.1.6 Implement and extract data for 2 behaviors . . . . . 5

1.1.7 Update documents . . . . . 6

1.2 Discussion of team member contribution . . . . . 6

1.2.1 Shreyas . . . . . 6

1.2.2 Shiru . . . . . 6

1.2.3 Yaqeen . . . . . 6

1.2.4 Mohammed . . . . . 6

**2 Plan for the next milestone** **7**

2.1 Discussion of each task . . . . . 7

2.1.1 Finish extracting information using the behaviors . . . . . 7

2.1.2 Begin Correletion of behavior information with grades . . . . . 7

2.1.3 Present information . . . . . 7

2.1.4 Design and Implement GUI . . . . . 8

**3 Sponsor feedback on each task for current milestone** **9**

**4 Sponsor Evaluation** **10**

# 1 Progress of current Milestone

| Task | Completed | Shreyas | Yaqeen | Shiru | Mohammed | Remarks |
|---|---|---|---|---|---|---|
| Finalize conceptualizing 8 behaviors | 100% | 28% | 35% | 22% | 15% | 14 Behaviours were identified |
| Finalize packages, languages and tools | 90% | 80% | 20% | - | - | Still need to examine the possibility of WEKA |
| Implement and test framework to get input from syllabus | 50% | - | - | 50% | 50% | More details from the syllabus need to be input. More exception handling |
| Select data mining techniques for each behavior | 80% | 33% | - | 33% | 33% | Some behaviors will have their method determined based on implementation |
| Provide report explaining behaviours and corresponding data mining methods | 100% | 25% | 25% | 25% | 25% | N/A |
| Implement and extract data for 2 behaviors | 100% | 100% | - | - | - | 3 behaviours implemented and extracted |
| Begin Preparing dataset | 80% | 100% | - | - | - | Need to use Regex/pyEnchant to get rid of non human-readable strings |
| Update Requirements Documents | 100% | - | - | 50% | 50% | N/A |
| Progress Evaluation | 100% | 40% | 20% | 20% | 20% | N/A |

## 1.1 Discussion of each task

### 1.1.1 Finalize Conceptualizing 8 behaviors

This task included examining the data to find behaviors that would aid in correlation to the grades. The team conceptualized 14 behaviors as opposed to the planned 8:

1. Number of days for each student with 0 activity[1]
2. Total number of Logins throughout the semester[2]
3. Average activities for each login [3]
4. Time between a due date and the first time relevant course material is accessed
5. Average weekly Logged in time
6. Number of assignments submitted on time
7. Average review time before assignments
8. Average review time before tests
9. Number of optional assignments done
10. Total time accessing study guides
11. Total time accessing podcasts
12. Frequency of accessing study guides
13. Frequency of Quiz reviews
14. Average time taken for quizzes

### 1.1.2 Finalize packages, languages and tools

The primary programming language was agreed upon to be **Python**, due to the high availability and quality of available open source data mining tools and libraries. Python also provides great API and framework to handle different file types and to present data The python libraries decided and implemented are:

- numpy - used to handle numbers and distributions, very widely and generically used library
- scipy - used to more advanced distributions, moments and kurtosis.
- scikit-learn - open source data mining/ machine learning library

[1] Implemented
[2] Implemented
[3] Implemented

- statsmodels
- pandas - makes working with tabular data like csv files very simple
- matplotlib - presentation
- pyEnchant - To distinguish plain english from gibberish and encoded strings
- sys - command line option parsing

Weka is being examined as a prospective tool

### 1.1.3 Implement and test framework to get input from syllabus

Simple CLI was built using python. To get all the surface data. This will be further detailed with having more parameters and better exception handling as well as data validation.

### 1.1.4 Select data mining techniques for each behavior

`https://github.com/shreyasugemuge/Senior-Design-Project/tree/master/docs/Milestone%202/Behaviors` comprises of four files with each group members research and analysis. 3 have been implemented, another layer of refinement is expected. Multiple Data mining methods for a single behavior may be used in order to find the best metric.

### 1.1.5 Clean and prepare data

The repository has been set up in a way that allows for the data to be on the local computer but ignored while being pushed. This enables a local working directory compatible to the scripts and programs written. The program CleanDataset.py serves to create a more program friendly log file. This will be refined if required. One thing that will be addded is filtering for encoded strings so the sample size of activities can be limited and hence quantified.

Certain Issues were identified and dealt with in this part, for instance the column containing quiz information also had duplications of all the object information. These were removed. The column titles were changed along with filenames for easier access.

A supervised Item based collaborative filtering would suggest that 0 as an activity is actually an outlier for the data And sessions that login and abandon the session should be treated as erronous data. This information was preserved however in order to treat them as a metric for behavior

### 1.1.6 Implement and extract data for 2 behaviors

3 behaviors were implemented as mentioned in 1.1.1. This subsection will explain the source directory structure and the programs.

- src/Behaviors
  - run.sh run the program in correct order and provides a verbose output while doing so
  - Behavior1.sh runs the beh_1.py to extract 3 behaviours
  - CleanDataset.sh run CleanDataset.py to clean all log files

#### 1.1.7 Update documents

Requirements document was updated, this will continue for the next milestone. The progress evaluation was drafted.

### 1.2 Discussion of team member contribution

#### 1.2.1 Shreyas

Conceptualized 4 Behaviors. Implemented 3 Behaviors. Finalized the python end of the program, including libraries. Wrote python script to prepare dataset. Updated website. Explored and dismiss possibility of unsupervised learning models. Explored possibility of of user based collaborative filtering using the given survey data, to further classify results. Update documents.

#### 1.2.2 Shiru

Conceptualized 2 behaviors, provided a report. Contributed to making the program to get syllabus information from the user. Contributed in updating the requirements document as well as progress evaluation.

#### 1.2.3 Yaqeen

Conceptualized 5 behaviors along with data mining technique. Provided a report. Contributed in updating progress evaluation.

#### 1.2.4 Mohammed

Conceptualized 2 behaviors along with data mining techniques, provided a report. Contributed to making the program to get syllabus information from the user. Contributed in updating the requirements document as well as progress evaluation

# 2 Plan for the next milestone

| Task | Shreyas | Yaqeen | Shiru | Mohammed |
|---|---|---|---|---|
| Finish extracting information using the behaviors (16 more behaviors to be implemented) | 25% | 25% | 25% | 25% |
| Begin Correletion of behavior information with grades | 25% | 25% | 25% | 25% |
| Present information as plots, graphs and charts | 25% | 25% | 25% | 25% |
| Design and begin implementing GUI. Must finish The design and UI side. | 25% | 25% | 25% | 25% |

## 2.1 Discussion of each task

### 2.1.1 Finish extracting information using the behaviors

The plan is to conceptualize 16 more behaviors for a total of 30. All of the implementations and information will not be used for the correlation. Decision trees to find the least entropy and lest chances of overfitting will qualify data, which is why having an abundance of behaviour will help.

### 2.1.2 Begin Correletion of behavior information with grades

In this part correlations between grades and behaviors will start being computed. Thresholds for confidence intervals will be determined.

### 2.1.3 Present information

matplotlib and statsmodels will help generate a good pictorial representation of the data and information we have. The behavior analysis will yield abstract human comprehensible reports for each student and for the class as a whole.

### 2.1.4 Design and Implement GUI

The Teacher end and Student end must be designed and clearly distinguished, with proper channels of information management to ensure product security. GUI will be mainly buttons, fields and check boxes for an intuitive yet easy communication.

# 3 Sponsor feedback on each task for current milestone

# 4 Sponsor Evaluation

Sponsor: Please detach this page and return to Dr. Shoaff

Score (0-10) for each member: circle a score (or circle two adjacent scores for .25 or write down a real number between 0 and 10)

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Shreyas Ugemuge | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 |
| Yaqeen AlKathiri | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 |
| Mohammed AlHabsi | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 |
| Shiru Hou | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 |

Faculty Sponsor

_____

*Signature*

_____

*Date*