```
\documentclass[12pt]{article}
\newcommand{\namesigdate}[2][5cm]{%
  \begin{tabular}{@{}p{#1}@{}}
    #2 \\[2\normalbaselineskip] \hrule \\[0pt]
    {\small \textit{Signature}} \\ [2\normalbaselineskip] \hrule \
\[0pt]
    {\small \textit{Date}}
  \end{tabular}
}
\usepackage{adjustbox}
\usepackage{color}
\usepackage{tabularx}
\usepackage{authblk}
\usepackage{tabto}
\usepackage{tcolorbox}
\usepackage{setspace}
\usepackage{listings}
\lstset
{ language=bash,firstnumber=1,numbers=left,numbersep=-10pt,basicstyle=
\ttfamily,
  showstringspaces=false,
  commentstyle=\color{red},
  keywordstyle=\color{blue}}
\usepackage[document]{ragged2e}
\usepackage[left=1in,right=1in,top=1in,bottom=1in]{geometry}
\usepackage[pdftex,pdfpagelabels,bookmarks,hyperindex,hyperfigures]
{hyperref}
\usepackage{graphicx}
\begin{document}
        \title{\textbf{Milestone 2 Progress Evaluation} \\ \hfill
\break
        Academic Behavior Reccomendation System}
        \author{Shreyas Ugemuge\      \texttt{sugemuge2014@my.fit.edu}
  \and
  Yaqeen AlKathiri\      \texttt{yalkathiri2013@my.fit.edu}
  \and
        Mohammed AlHabsi\      \texttt{malhabsi2013@my.fit.edu}
  \and
  Shiru Hou\      \texttt{shou2015@my.fit.edu}
  \and
  Faculty Sponsor: Dr. Phillip Chan\      \texttt{pkc@cs.fit.edu}}
        \maketitle
        \pagebreak
        \singlespacing
        \tableofcontents
        \pagebreak
        \section{Progress of current Milestone}
        \begin{adjustbox}{totalheight=\textheight-2\baselineskip}
        \begin{tabularx}{\textwidth}{|X|c|c|c|c|c|X|}
        \hline
```

```
                \textbf{Task} & \textbf{Completed} &\textbf{Shreyas} &
\textbf{Yaqeen} & \textbf{Shiru} & \textbf{Mohammed} &
\textbf{Remarks}  \\
                \hline
                Finalize conceptualizing 8 behaviors & 100\% & 28\% &
35\% & 22\% & 15\% & 14 Behaviours were identified \\ \hline
                Finalize packages, languages and tools & 90\% & 80\% &
20\% & - & - & Still need to examine the possibility of WEKA \\ \hline
                Implement and test framework to get input from
syllabus & 50\% & - & - & 50\% & 50\% & More details from the syllabus
need to be input. More exception handling\\ \hline
                Select data mining techniques for each behavior & 80\%
& 33\% & - & 33\% & 33\% & Some behaviors will have their method
determined based on implementation\\ \hline
                Provide report explaining behaviours and corresponding
data mining methods & 100\% &25\% & 25\% & 25\% & 25\% & N/A\\ \hline
                Implement and extract data for 2 behaviors & 100\%
&100\% & - & - & - & 3 behaviours implemented and extracted\\ \hline
                Begin Preparing dataset & 80\% & 100\% & - & - & - &
Need to use Regex/pyEnchant to get rid of non human-readable strings \
\  \hline
                Update Requirements Documents & 100\% & - & - & 50\% &
50\% & N/A \\ \hline
                Progress Evaluation & 100\% & 40\% & 20\% & 20\% & 20\
% & N/A \\ \hline
        \end{tabularx}
        \end{adjustbox}
        \subsection{Discussion of each task}
                \subsubsection{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:
                \begin{enumerate}
                        \item Number of days for each student with 0
activity\footnote{Implemented}
                        \item Total number of Logins throughout the
semester\footnote{Implemented}
                        \item Average activities for each login
\footnote{Implemented}
                        \item Time between a due date and the first
time relevant course material is accessed
                        \item Average weekly Logged in time
                        \item Number of assignments submitted on time
                        \item Average review time before assignments
                        \item Average review time before tests
                        \item Number of optional assignments done
                        \item Total time accessing study guides
                        \item Total time accessing podcasts
                        \item Frequency of accessing study guides
                        \item Frequency of Quiz reviews
```

```
                        \item Average time taken for quizzes
                \end{enumerate}

                \subsubsection{Finalize packages, languages and tools}
                The primary programming language was agreed upon to be
\textbf{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:

                \begin{itemize}
                        \item numpy - used to handle numbers and
distributions, very widely and generically used library
                        \item scipy - used to more advanced
distributions, moments and kurtosis.
                        \item scikit-learn - open source data mining/
machine learning library
                        \item statsmodels
                        \item pandas - makes working with tabular data
like csv files very simple
                        \item matplotlib - presentation
                        \item pyEnchant  - To distinguish plain
english from gibberish and encoded strings
                        \item sys - command line option parsing
                \end{itemize}
                Weka is being examined as a prospective tool

                \subsubsection{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.
                \subsubsection{Select data mining techniques for each
behavior}
                \url{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.
                \subsubsection{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.  \\ \hfil \break
```

```
                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. \\
\hfil \break
                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
                \subsubsection{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.
                \begin{itemize}
                        \item src/Behaviors
                        \begin{itemize}
                                \item run.sh run the program in
correct order and provides a verbose output while doing so
                                \item Behavior1.sh runs the beh\_1.py
to extract 3 behaviours
                                \item CleanDataset.sh run
CleanDataset.py to clean all log files
                        \end{itemize}
                \end{itemize}
                \subsubsection{Update documents}
                Requirements document was updated, this will continue
for the next milestone. The progress evaluation was drafted.

        \subsection{Discussion of team member contribution}
        \subsubsection{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.
        \subsubsection{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.
        \subsubsection{Yaqeen}
        Conceptualized 5 behaviors along with data mining technique.
Provided a report. Contributed in updating progress evaluation.
%Implemented one behavior using WEKA
        \subsubsection{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
        \section{Plan for the next milestone}
        \begin{tabularx}{\linewidth}{|X|X|X|X|X|}
        \hline
        \textbf{Task} & \textbf{Shreyas} & \textbf{Yaqeen} &
\textbf{Shiru} & \textbf{Mohammed} \\ \hline
        Finish extracting information using the behaviors (16 more
behaviors to be implemented) & 25\% & 25\% & 25\% & 25\% \\ \hline
        Begin Correletion of behavior information with grades & 25\% &
25\% & 25\% & 25\% \\ \hline
        Present information as plots, graphs and charts & 25\% & 25\%
& 25\% & 25\% \\ \hline
        Design and begin implementing GUI. Must finish The design and
UI side. & 25\% & 25\% & 25\% & 25\% \\ \hline
        \end{tabularx}

        \subsection{Discussion of each task}
        \subsubsection{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.
        \subsubsection{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.
        \subsubsection{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.
        \subsubsection{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.

        \pagebreak
        \section{Sponsor feedback on each task for current milestone}
        \pagebreak
        \section{Sponsor Evaluation}
        Sponsor: Please detach this page and return to Dr. Shoaff \\
\hfill \break
        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)
\\ \hfill \break
```

```
        \begin{tabularx}{\textwidth}{|X|c|c|c|c|c|c|c|c|c|c|c|c|
c|}
        \hline
        Shreyas Ugemuge & 0 & 1 &  2 & 3 & 4 & 5 & 6 & 6.5 & 7 & 7.5 &
8 & 8.5 & 9 & 9.5 & 10 \\ \hline
        Yaqeen AlKathiri & 0 & 1 &  2 & 3 & 4 & 5 & 6 & 6.5 & 7 & 7.5
& 8 & 8.5 & 9 & 9.5 & 10 \\ \hline
        Mohammed AlHabsi & 0 & 1 &  2 & 3 & 4 & 5 & 6 & 6.5 & 7 & 7.5
& 8 & 8.5 & 9 & 9.5 & 10 \\ \hline
        Shiru Hou & 0 & 1 &  2 & 3 & 4 & 5 & 6 & 6.5 & 7 & 7.5 & 8 &
8.5 & 9 & 9.5 & 10 \\
        \hline
        \end{tabularx}
        \hfil \break
        \hfil \break
        \namesigdate{Faculty Sponsor}
        \end{document}
```