Math 502B – Probability & Statistics II – Summer 2016 – Class Project
Statistical Methods in Cancer Detection Using DNA Microarray Data

Group	Name	Name	Name	Project Investigation
1	*Duy*	*Khoa*	*Alejandro*	*Penalized Logistic Regression(PLR) & Nearest Shrunken Centroids(NSC)*
2	*Karl*	*Nick*	*Anthony*	*Bagging-Boosting Classification Trees (BBCT) & Random Forest (RF)*
3	*Silvia*	*Alice*	*Luis*	*2-Stage Support Vector Machine – Recursive Feature Elimination (SVM-RFE)*
4	*Matthew*	*Jacob*		*Multi-Dimensional Ranking (MDR)*
5	*Nathan*	*Chrystal*		*Distribution-Based Classification (DBC)*
6	*Guikai*	*Mayra*	*Alejandra*	*Decision Tree (DT) & Naïve Bayesian (NB)*

Goals:

- Understand gene expressions in DNA microarray data
- Understand the assigned statistical methods in cancer detection
- Develop computational algorithms for cancer detection
- Assess the algorithm performance
- Analyze accuracy for different cancer sets
- Summarize findings
- Present an oral presentation and write a technical report

By This Date	*Milestone/Achievements*
Wed 07/06/2016	**Project Introduction, Assigned Methods, Data Sets**
Wed 07/13/2016	**Mathematical Equations for Classification Method(s)**
Wed 07/20/2016	**Numerical Implementation for Classification**
Wed 07/27/2016	**Case Studies Investigation & Abstract Due**
Wed 08/03/2016	**Project Presentation/Technical Paper Due**

Project outcomes: A technical report and a well-rehearsed presentation that addresses the following sections

a. **Title & Abstract** summarizing your presentation (~ 500 words submitted electronically **07/27/2016**)
b. Introduction & Motivation
   What is your project about?
   Why is it important?
   What are you trying to do?
   How your study/analysis impact a real life scenario, if any.
   Others
c. Statistical Method
   What does it describe?
   How was the method derived?
   The meanings of your variables and parameters (type, units, range, etc)
   Additional assumptions?
d. Solution Process:
   Describe the proposed work
   Address the software/code used in the simulation
   Verify your claims numerically
e. Your findings:
   Describe the findings
   Significant contributions your group have found
f. Conclusion:
   Your contributions
   Summarize your work and your findings and possible future work
g. Reference:
   Cite any sources that you have used, including internet findings.

2. **Computer Codes:** Computer codes used for the project with detailed description of what each code does including the description of all input and output variables. You need to provide a readme.txt file explaining the chronological procedure how to run the codes.
3. **Grade:** Effort & Progress (10%) Oral Presentation (10%) Written Technical Report (10%) = Total (30%)