Mini-Project #3 (censored)

	Your Score	Problematic	OK	Good	Excellent	
Analysis & Exposition						
Introduction		0.4	0.6	0.8	1	Introduction describes the nature of the problem with thoroughness. The problem is framed in such a way that the topic is compelling.
Objective #1♦: Clustering the texts in several ways (Step 1)		0.5	1.5	2.5	3	Data are standardized before analysis. A variety of different clustering approaches are clearly described and used. [CENSORED]
Objective #2♦: Using classification analysis (aka discriminant analysis) (Step 2)		0.5	1.5	2.5	3	When classification analyses are used to compare cluster distinctness: (i) cross-validation is properly described and used, (ii) [CENSORED], (iii) multiple methods are used for creating classification rules and a justifiable approach is used for choosing the method chosen (e.g., linear vs. quadratic vs. KNN)
Objective #3♦: Assessing cluster distinctness and selecting a clustering for further analysis (Step 2)		0.5	1.5	2.5	3	Developed a coherent approach for characterizing the utility of competing clusterings. Each of the following are addressed: misclassification rates from a classification analysis, MANOVA tests of cluster mean equality, and the relative sizes of the clusters. The choice of which clustering to use is based on more than a simple comparison of misclassification rates (e.g., [CENSORED]).
Objective #4♦: Characterization of the "natural/data-based clusters" from the selected clustering (Step 3)		0.5	1	1.5	2	Discriminant analysis is properly employed to characterize the separation of the clusters in s-space, noting their relative collinearity or diffuseness; [CENSORED]. Each cluster is characterized and labeled, using discriminant function coefficients, cluster means, and/or genre labels.
Objective #5♦: Comparing clusters with genres (Step 3)		0.5	1	1.5	2	A two-way table (or comparable device) is used to show how the clusters in the chosen clustering align with the 15 genres. The best 5-cluster solution is compared with the super-genre variable. The relative strengths and weaknesses of the natural 5 clusters and the 5 super-genres are discussed (e.g., [CENSORED]). [CENSORED].
Conclusion		0.4	0.6	0.8	1	Important results are briefly summarized, and limitations and future work are described.
Document Quality						
Figures & Tables		0.5	1	1.5	2	Tables and Figures are placed in the document as Latex "floats" that appear at the tops of pages. Captions are appropriate (above each table and below each figure) and allow the reader to understand the structure of the table/figure. All tables and figures are discussed/analyzed in the text with references made to the table/figure of interest; important interpretations and insights gained are explicitly stated in the text.
References		0.3	0.3	0.3	0.5	Appropriate references are given.
Typos & Formatting		0.6	0.9	1.2	1.5	Document is free of typos and formatting is clean and professional.
Well-commented code		0	0.5	0.8	1	Code is complete and well-commented.

♦ For full credit on project objectives, the following will be expected in a thorough coverage of the objective:

- ***Technically Thorough and Intuitive Explanations***: Ideas that are new to the target audience (i.e., the client) are explained using both a technical definition (usually a formula or sketch of an algorithm) *and* an intuitive description.
- ***Numerical Results***: Numerical results are made available to the reader and such results are described in the text and/or formatted in tables to optimize the exposition. E.g., discriminant function coefficients are both interpreted provided in a table, estimated Box-Cox values are both interpreted and reported in a table, p-values are explicitly reported, etc.