Dr. Eick

COSC 6335 *“Data Mining”* Fall 2011

Draft Project5: Learning and Assessing Classification Models

Individual Assignment

Due: Friday, December 2, 11p[[1]](#footnote-1) (electronic Submission)

Last Updated: November 9, 2011, 2p

The goal of the project is the development and assessment of classification models for the Vehicle Silhouettes Data Set dataset from [http://archive.ics.uci.edu/ml/datasets/Statlog+(Vehicle+Silhouettes)](http://archive.ics.uci.edu/ml/datasets/Statlog%2B%28Vehicle%2BSilhouettes%29). Students should

1. use a single classification approach[[2]](#footnote-2), such as K-NN, decision trees,… to obtain classification models for the Vehicle Silhouettes Data Set. Preferably, us R classification tools.
2. conduct some (limited) experiments that analyze the impact of different classification tool parameters on the accuracy of the obtained model.
3. interpret the project results.
4. summarize your finding in a 2-3 page single-spaced report
5. submit other deliverables as appendixes of the report, if applicable, such as:
	1. Best model found
	2. New distance functions created
	3. Source code of software you developed
	4. Other outcome which you believe shed light on the approach you used in the project.
1. In general, you should be able to finish the project by November 28, but some of you might be out of town over Thanksgiving which is the justification for using a much later deadline! [↑](#footnote-ref-1)
2. Any approach is fine! [↑](#footnote-ref-2)