

**Your Name:**

**Your Andrew ID:**

## **Homework 2**

### **Collaboration and Originality**

1. Did you receive help of any kind from anyone in obtaining your data for this assignment (Yes or No)?  
It is not necessary to describe discussions with the instructor or TA.

If you answered Yes, provide the name(s) of anyone who provided help, and describe the type of help that you received.

2. Did you give help of any kind to anyone in obtaining their data for this assignment (Yes or No)?

If you answered Yes, provide the name(s) of anyone that you helped, and describe the type of help that you provided.

3. Are you the author of every word of your report (Yes or No)?

If you answered No:

- a. identify the text that you did not write,
- b. explain where it came from, and
- c. explain why you used it.







	$n_1=00$			$n_2=00$			$n_3=00$			$n_4=00$			$n_5=00$		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
<b>J48</b>	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
<b>NB</b>															
<b>SVM</b>															

## 2.2 Analysis

Discuss your results. Are small representations more effective, or are large representations more effective? Does each dataset and/or learning algorithm behave the same way? Does having more features make a difference? If so, is it an important difference? Are the effects similar for small and large classes?

## 3 Experiment #3: Your Representation

Report your results in a tabular format. Describe your representations. Discuss your reasons for developing each representation, and how well they worked.

## 4 Effectiveness of the Different Representational Choices

Do you see any trends? Do some techniques consistently work better than others? Do any of the choices affect the stability of results? Do any of the differences seem significant?

There is some overlap between this section and observations that you may make in Experiments 1-3. However, in this section you have an opportunity comment on trends that span different experiments, which you can't really do in the earlier sections.

## 5 Effectiveness of the Different Learning Algorithms

Do you see any trends? Does one algorithm consistently work better than another? Is there a difference in the stability of the algorithms? Do any of the differences seem significant?

There is some overlap between this section and observations that you may make in Experiments 1-3. However, in this section you have an opportunity comment on trends that span different experiments, which you can't really do in the earlier sections.

## 6 Effectiveness on Small, Medium, and Big Categories

Are categories with more training data easier to learn? If so, is it an important difference?

## 7 Decision Tree Categories

Pick one of the datasets. Pick 2 categories that the decision tree algorithm did well on, and 2 categories that it didn't do well on. Examine **the first few levels of the decision tree**. Develop **human-readable** rules for each of these categories? (**Do not** show the entire tree. Recall the simple example covered in

class involving "cents" and "vs".) Does this help you understand why the decision tree did well on some categories, but not others?