Predictive Model on "Fake Reviews"

BY: BASSAM KAAKI

COURSE: MANAGEMENT 647

Agenda

- Introduction
- What is the problem?
- Web Page link & selected product
- Approach
- Classification of messages
- Data sheet
- Predictive model (3 models)
- Predictive model first type (Class)
- Decision tree rules (Class)
- Predictive model 2nd type (Class without reviews)
- Decision tree rules (Class without reviews)
- Predictive model 3rd types (Class with reviews)

- Decision tree rules (Class with reviews)
- Comment examples
- Comparing class with and without reviews
- Suggestions for online retailors to avoid fake reviews

Introduction

- Customers look at online reviews for information on a product
  - Compare prices
  - Search for functionality and durability
  - Obtain knowledge to make purchase decisions
- Online retailors are plagued with fake online reviews
- How do online businesses such as Google, Amazon or Yelp deal with this?
  - Algorithms (no clear spamming patterns) that flag reviews/comments
  - Less than 80% effective

What is the problem?

- A model that will identify fake reviews on an Amazon page
  - Quantify probability
  - Identification of important predictors
- What are we looking for in the model
  - Accuracy
  - Easy to understand
  - Easy to implement
  - Robust

Web Page Link and selected product

- Amaxon page (Scarleton Large Drawstring Handbag H1078)

- Link: https://www.amazon.com/Scarleton-Large-Drawstring-Handbag-H107804/dp/B009DKKUP8/ref=sr_1_sc_3?ie=UTF8&qid=1519675780&sr=8-3-spell&keywords=scaleton+bag

Approach

- **Data**
  - First step:
    - Google scraper
      - Extract reviews information
  - Second Step:
    - Data classification of 1,000 reviews
      - 13 fields based on size, smell, color, like it, gift, quality, problem…
      - 1 = present in message 0 = not present in message
      - Class contains whether message is False/True based on human sentiment
  - Third Step:
    - Model Creation

Classification of messages

- Very descriptive comments using "height", "width", "texture name"
- Marketing speak, "selling tone", "pushing to buy tone"
- Generic comments such as "like it", "love it", "adore it", "nice", "excellent"
- Grammatical errors in the comments
- Comments does not fully explain why a person like or dislikes product
- Helpful comments
- Overuse of product name
- Product relations
- Same ID's commenting more than once
- Giving 5 star ratings

**Customer Reviews**

Data Sheet

ID	Reviews	Helpful comm	Short messa	Long messa	Qual	Si	Sm	Co	Lov	Like	Probl	G	Deliv	Cl
15207589512352	Well pleased with this handbag. … it's good quality. … just the right size and the color is awesome. ….	0	1	0	1	1	0	1	0	0	0	0	0	T
15207587381520	Great purse! I have gotten many compliments and it has lasted longer than any of my other purses!	0	1	0	0	0	0	0	0	0	0	0	0	T
15207586691202	It was a little too large for me, but nice. The color was a true blue not navy. Nice style	0	1	0	0	0	0	1	0	0	0	0	0	T
15207586521128	Bought this bag for my daughter. Looks great. You can't tell it's not leather. Picked out a brown one and she loved the color. I liked the bag so much I ordered a burgundy one.	0	1	0	1	0	0	1	0	0	0	1	0	T
15207590982882	It has a lot of space for those of us who take our whole house in our purse. I use this as my go to purse for my lengthy work commute. I can even stick a small lunch bag in there. I can't decide if it looks cheap or not but it was certainly cheap haha. Anyway, I wanted something that wasn't leather and could take a beating on the bus/shuttle rides I use to commute and it has definitely held up.	0	0	1	0	0	0	0	0	0	0	0	0	T
15207589272256	This purse was exactly what I wanted. It is a large bag that can fit all my things and yet comfy to wear when it gets heavy. Good amount of pockets and it looks great! Lots of compliments. Love this bag!	0	1	0	0	0	0	0	0	0	0	0	0	F
1520758618980	Really beautiful, stylish bag! Just love!!	0	1	0	0	0	0	0	0	0	0	0	0	F
15207589272257	Doesn't really look like real leather, but it's cute. Maybe I expected too much. I'll use it but don't plan on buying any more. Save your money and go to designer handbag outlet stores	0	1	0	0	0	0	0	0	1	0	0	0	T
15207589912503	This is my first bag by Scarleton and it definitely won't be my last! I will be ordering more colors of this bag. I have another, different style, bag by Scarleton already ordered, but I haven't received it yet. I can tell it will be a winner though. If you didn't know what this bag was made from, you would be hard pressed to say it wasn't leather. The material is sturdy and thick and substantial. The hardware is really pretty and looks like it will last for a very long time. The zippers work smoothly. The bag is the absolute perfect size. Not too huge, but certainly not small by any means. I love the crossbody strap that allows me to carry the bag hands free. This is my new favorite brand!	0	0	1	0	1	0	1	1	0	0	0	0	T
15207587791701	Recommend!	0	1	0	0	0	0	0	0	0	0	0	0	F

Predictive models (3 types)

1. **Class** – model prediction using human sentiment based on actual data
2. **Class without reviews** – model prediction using human sentiment, no comments categorization and based on actual data (training and testing sets)
3. **Class with reviews** – model prediction based on human sentiment from actual data including categorized comments (training and testing sets)
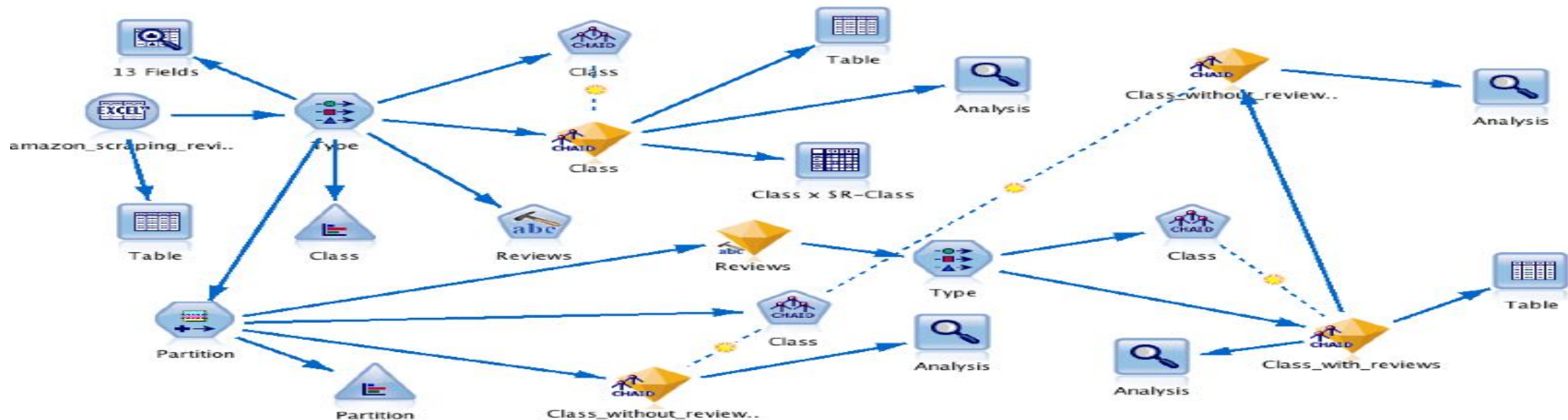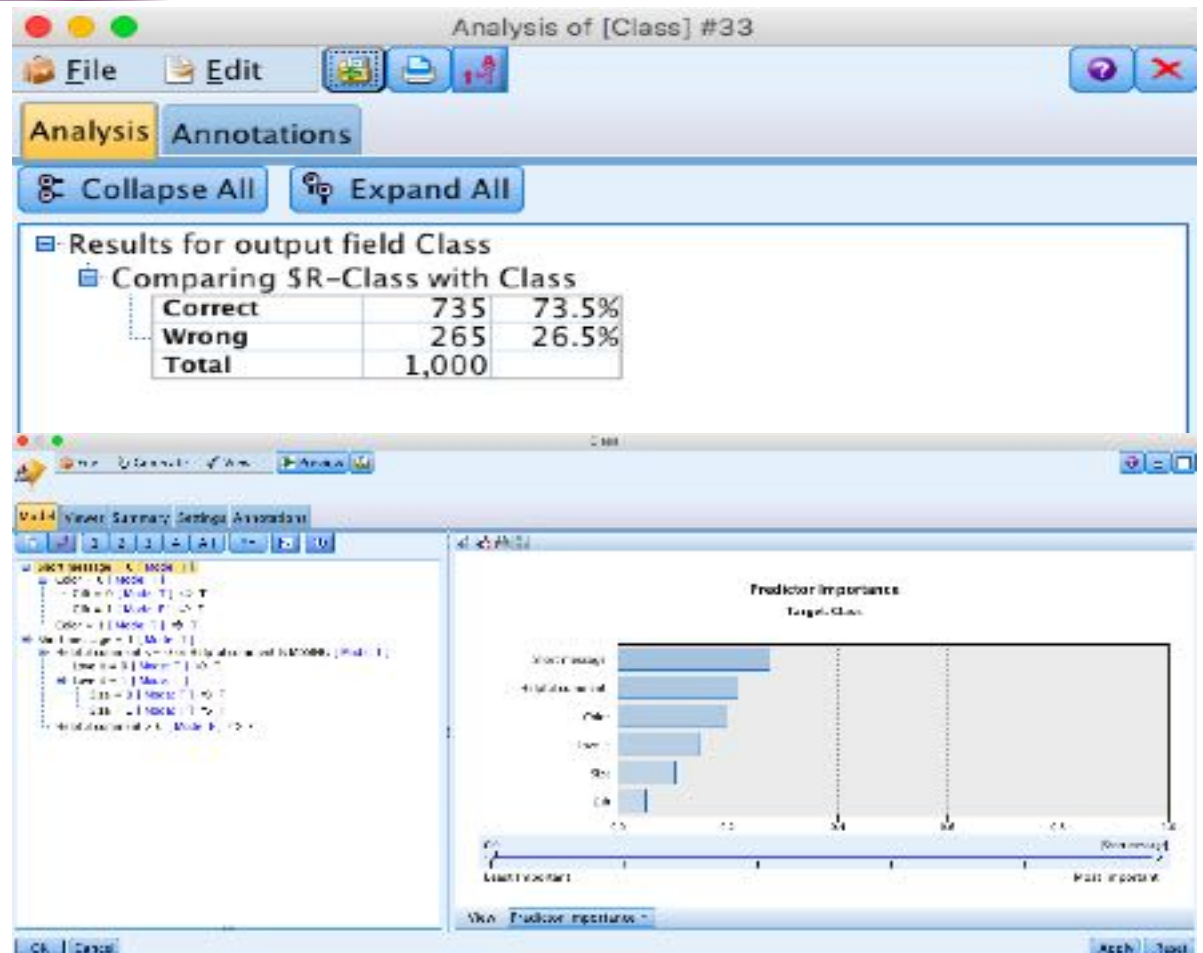
Predictive model 1ˢᵗ type (Class)

- Class (100% raw data)
- 13 fields used
- Comparison of model prediction with my actual prediction for comments is 73.5% correct while 26.5% were wrong

*wrong means: SPSS classified the true as False and vice versa.

- Most important predictors
  - Short comment (28%)
  - Helpful comment (22%)
  - Color (20%)

Decision Tree Rules (Class)

- If comment is not short, does not contain the word color and the word gift, then its true.
- If comment is not short and does not contain the word color but contains the word gift then it is most likely false.
- If comment is not short and contains the word color then its true.
- Short comment, no helpful comment, comment is missing, no love word, is a true comment
- Short comment, no helpful comment, comment is missing, contains love word and no size word comment is true
- Comment is short, no helpful comment, comment is missing, love and size word included then comment is false

Predictive model 2nd Type (class without reviews)

- Class without reviews (Partitioned 79.90% ; 20.01%)

- 13 fields used

- Comparison of model prediction with my actual prediction for comments is 73.34% correct & 26.66% wrong

- Validating the data with the 20.01% yielded 72.14% correct and 27.86% wrong

- Most important predictors

  - Gift (67%)

  - Short message (19%)

  - Color (14%)
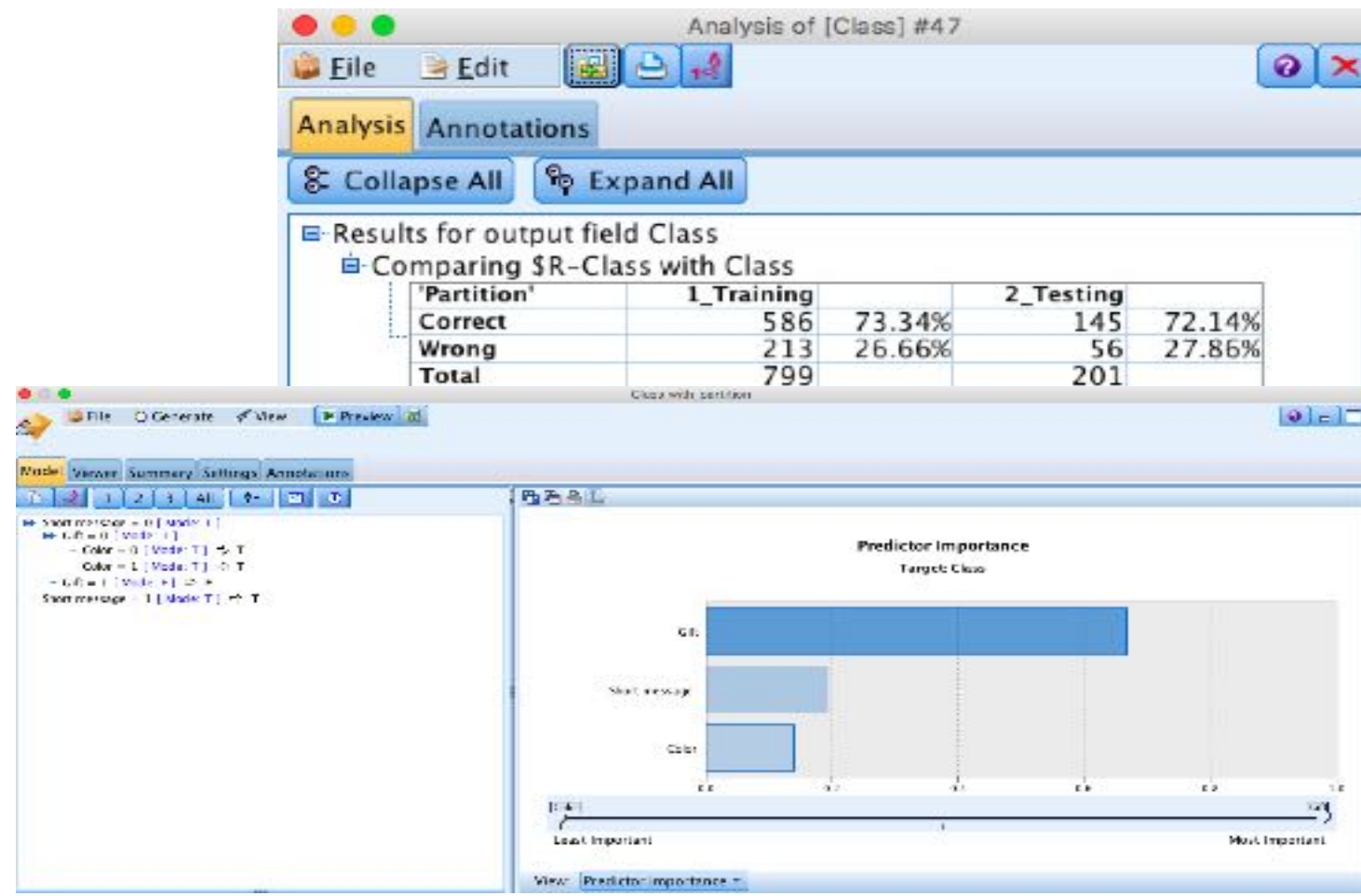
Decision Tree Rules (class without reviews)

► If its not a short comment, does not contain the word gift or the word color then review is true.

► If its not a short comment, does not contain the word gift, but contains the word color, then its true.

► If its not a short comment, contains the word gift, then its false.

► Short comments are considered true.

Predictive model 3rd Type (Class with reviews)

- Class with reviews (Partitioned 79.90% ; 20.01%)
- Comparison of model prediction with my
actual prediction for comments is 73.34% correct
& 25.66% wrong
- Validating the data with the
 20.01% yielded 73.13% correct and
26.87% wrong

- Most important predictors
  - Consumer category (18%)
  - Pockets (17%)
  - Makeup sub-category smell (14%)
  - Product attributes (12%)

⊟ Results for output field Class
  ⊟ Comparing $R–Class with Class
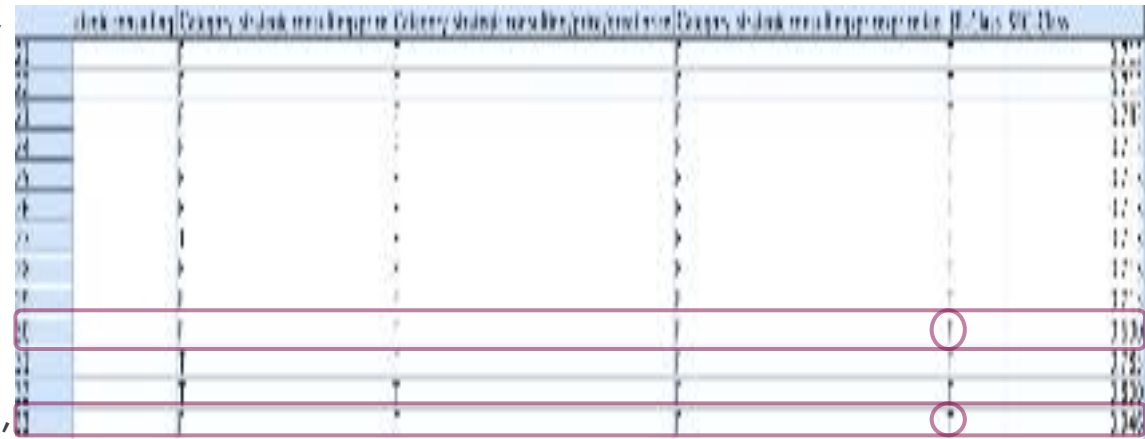
'Partition'	1_Training		2_Testing	
Correct	594	74.34%	147	73.13%
Wrong	205	25.66%	54	26.87%
Total	799		201	
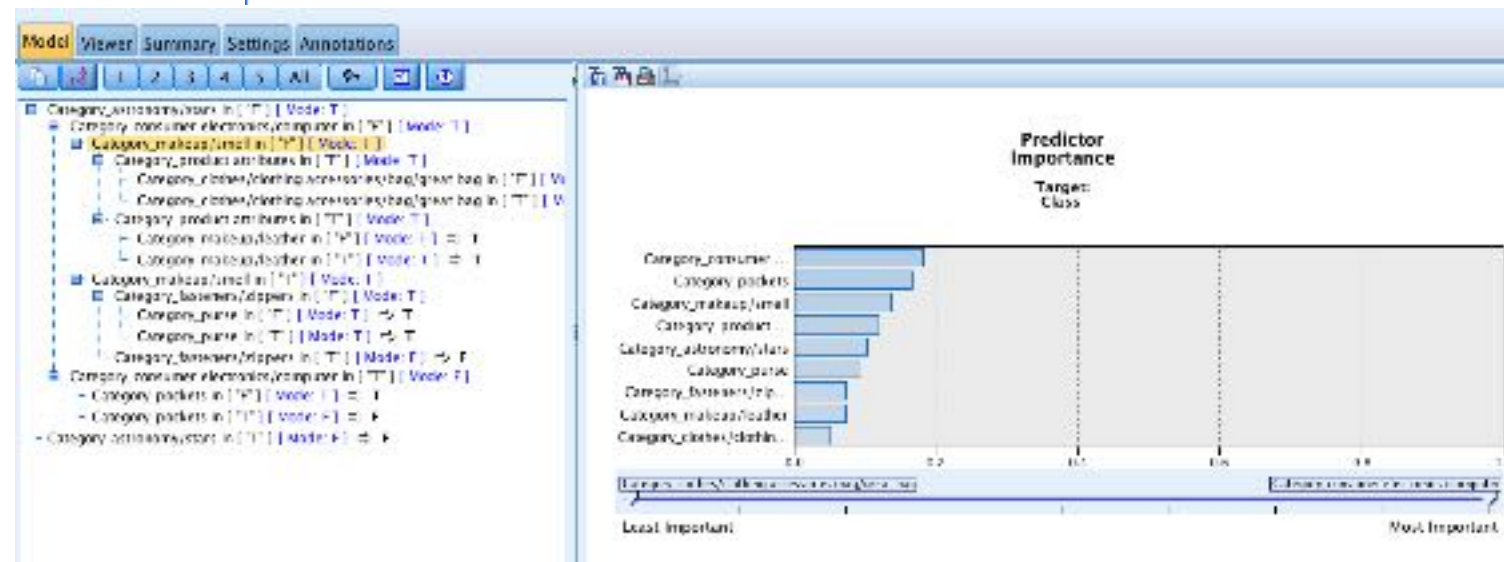
Decision tree rules (class with reviews)

- A comment that is classified under consumer (18%) predictor importance, contain words such as gold, hardware, item, mobile, phone, smartphone, computer, tablet, kindle, body strap, color, nice color, consumer, fasteners, look, pockets, accessories.

- Model prediction points out that these comments are likely to be false with a confidence level of **60%**.

- If comments in the consumer category (18%) predictor importance contained words such as clothes, handbag, bag, finance, purchase, room, odor, smell, feel of leather, leather, makeup, then these messages are likely to be true.

- Model prediction for these comments are **94.6%** true.

Comment examples

- Comment (no:30) **FAKE**

  - "The only reason I gave this a 4-star rating was because of the cross-body strap. I wish it was wide as the shoulder strap and perhaps thicker considering it drags more when you just let it hang, especially since one of the main reasons I'll be using that feature is when I am carrying heavier items. The straps just ends up digging into my shoulder uncomfortably. Otherwise, I love the size, the pockets, the look and since the bag I bought was sea green, the gold hardware blended very well for an overall nice coloring. No regrets, would purchase again, especially if they fix the width of the cross-body strap"
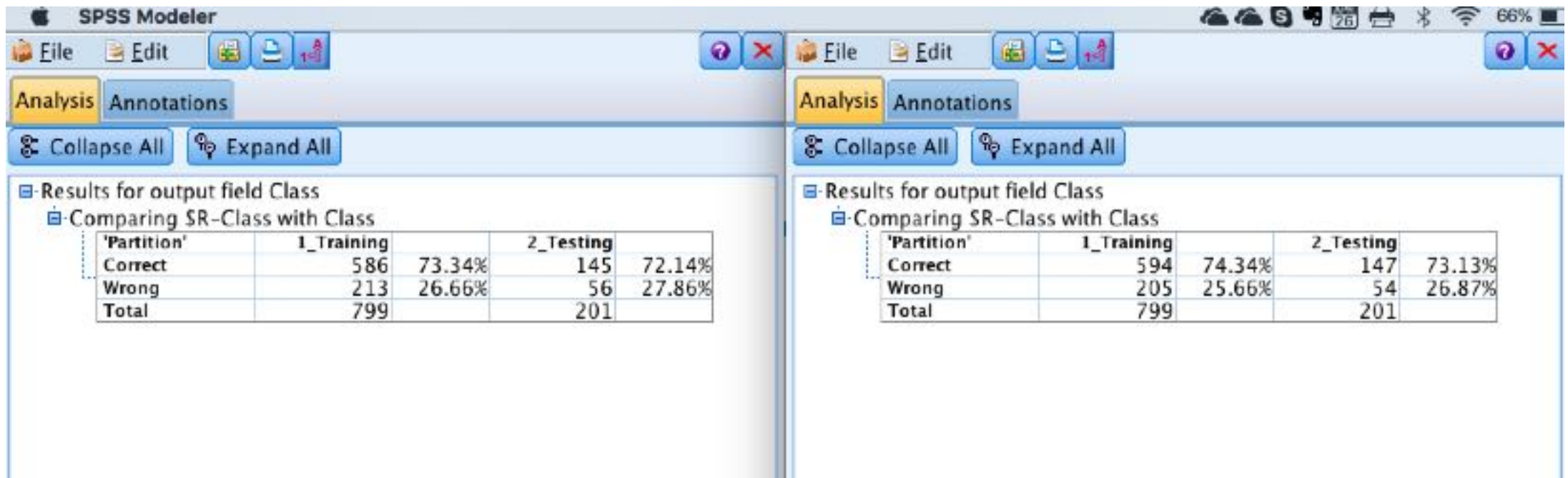
- Comment (no:33) **TRUE**

  - " I really like my new handbag it is large enough to carry everything I need plus I have more room than what I thought I would. It has a very soft leather feel to it. Only drawback is the inside has a odor. Before I bought his handbag, I read some of the remarks and saw that was a problem with a few people. I thought I would buy it anyway. So far, I really do like my new purchase. I have had this for a week and hoping the smell will eventually come out soon"

Comparing class with and without review results

▶ Success rate when categorizing the comments increased to 1% in predicting the model using the training set and 0.99% using the validation set.

▶ The 1% success rate, although small is a good indication that this model can be scaled to larger data sets.

Suggestions for online retailors to avoid fake reviews

- Add mood tickers
- Connect certified purchaser with reviewer
- Anonymous users not allowed to write review
- Comment level (dynamic graphic)
- Rules/policies – place reviewer liable
- Voting value - connect comments/reviews that affected customer purchase

Thank you!

QUESTIONS?