BAN 5743
Exercise 7 (10 Points)
Sentiment Analysis Solution

In this exercise, you will use SAS Sentiment Analysis Studio with positive and negative movie reviews collected from IMDB.com.

Data:

Part I:

The data for *modeling* are stored in these folders:

- Positive reviews : Movie_Review_Sentiments\Model_data\Positive
- Negative reviews: Movie_Review_Sentiments\Model_data\Negative

The data for *testing* are stored in these folders:

- Positive reviews : Movie_Review_Sentiments\ Testing_data\Positive
- Negative reviews: Movie_Review_Sentiments\Testing_data\Negative

Part II:

The data for *modeling* are stored in these folders:

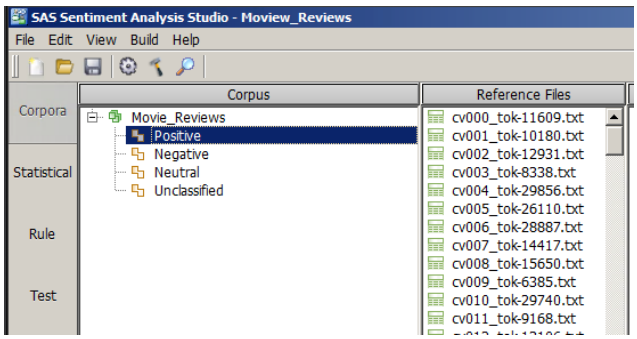
- Positive feedbacks : Sentiment_Mining_data\Model_data\Positive
- Negative feedbacks: Sentiment_Mining_data\Model_data\Negative

The data for *testing* are stored in these folders:

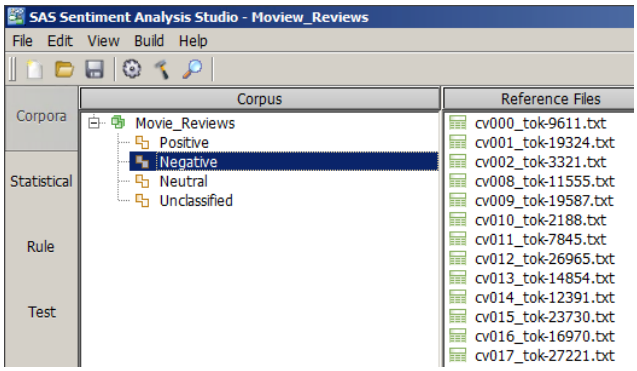
- Positive feedbacks : Sentiment_Mining_data\ Testing_data\Positive
- Negative feedbacks: Sentiment_Mining_data\Testing_data\Negative

Questions - Part I:

- ✓ Create a new project in Sentiment Analysis Studio using the default settings for rule-based and statistical models.
- ✓ Click the **Corpora** tab. Right-click in the white space and select a new corpus. Give it a name such as **Movie_Review**.
- ✓ Right-click **Positive** under Movie_Review and select **Add a directory**. Select **Movie_Review_Sentiments ⇨ Model_data ⇨ Positive**.



✓ Right-click **Negative** under Movie_Review and select **Add a directory**. Select **Movie_Review_Sentiments ⇨ Model_data ⇨ Negative**.



✓ Select **Statistical**. Right-click in the white space and select **New Model**. Create a new statistical model named Part1 with advanced settings by clicking **Advanced ⇨ OK** in the Add New Model window.

✓ Build the Movie_Review_Adv model by selecting **Build ⇨ Build Statistical Model**.

1. **Report** the results and **comment** on it. **(1 point)**

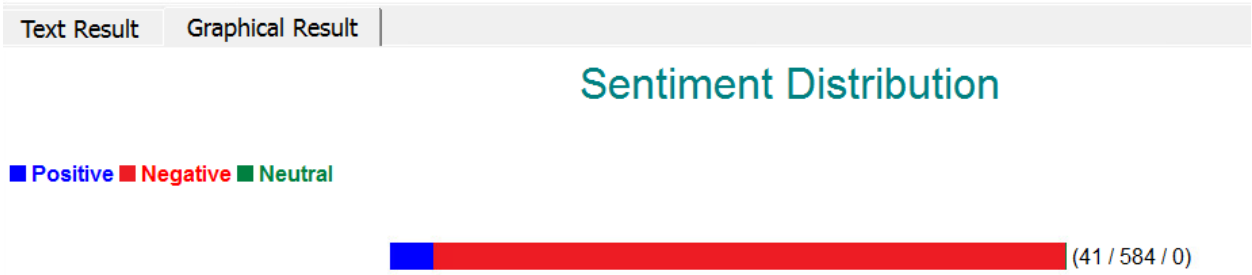
The best model was the Smoothed Relative Frequency and Information Gain. It performed the best on positive, negative and overall sentiments.

✓ Click **Test,** right-click in the white space of Test Data, and select **New Test Directory**. Point to the negative folder in the testing data directory.

✓ Right-click the folder under Manual Test and select **Test in Statistical Model**.

2. **Report** the results and **comment** on it. **(1 point)**

Since the negative folder location was tested, we would expect the results of the analysis to be mostly negative in sentiment. 93% of the reviews were correctly classified as negative (584/625).



✓ Repeat the testing steps with the positive test directory.

3. **Report** the results and **comment** on it. **(1 point)**

Since the positive folder location was tested, we would expect the results of the analysis to be mostly positive in sentiment. 98% of the reviews were correctly classified as positive (615/625).

✓ Click the **Rule** tab. From the top menu, select **Build ⇨ import learned features**.

✓ Select the statistical model created in previous steps – Part1.

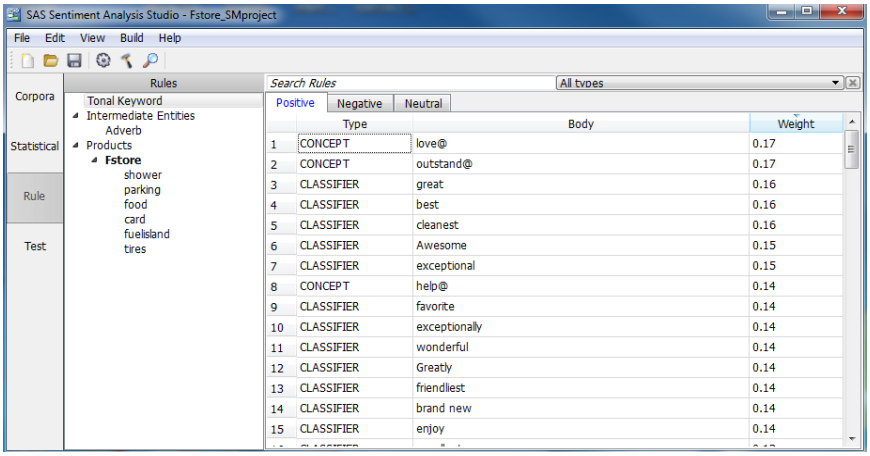
✓ Select **Tonal Keyword** and then click the **Positive** or **Negative** tab to get a sense of the rules.

4. **Comment** on the results. Do you see any Classifiers that do not make sense? **(1 point)**

There are some classifiers that do not seem readily positive or negative included in the lists. For example, "jessica" is included as a negative classifier and "spoil" is included as a positive classifier. The majority of classifiers do make sense though.

Questions - Part II:

- ✓ Create a new project in Sentiment Analysis Studio using the default settings for rule-based and statistical models.
- ✓ Click the **Corpora** tab. Right-click in the white space and select a new corpus. Give it a name such as **Fstore**.
- ✓ Right-click **Positive** under Fstore and select **Add a directory.** Select **Sentiment_Mining_data ⇨ Model_data ⇨ Positive**.
- ✓ Right-click **Negative** under Fstore and select **Add a directory**. Select **Sentiment_Mining_data ⇨ Model_data ⇨ Negative**
- ✓ Select **File ⇨ Import Rules** to import **Fstore_CustomRule.xml**.



- ✓ Click and explore some of the rules to get a sense of what these rules are and how they function.

1. Can you explain the Positive tonal keyword rule for Thank@? **(1 point)**

   As long as morphological expansion is enabled (default setting), the @ sign is representative of any item that would be an expansion of Thank such as thanks, thanking, thanked.

2. Can you explain the negative Concept_Rule for FStore_Shower? **(1 point)**

   The negative concept rule for FStore_Shower indicates that the entity "shower" in conjunction with the listed words such as dirty, hair, or cold indicate a negative tone in the document. All the negative concept rule for Fstore shower are listed below

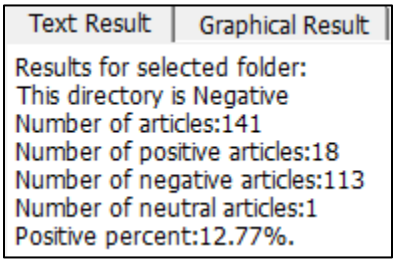
Definitions | Positive | Negative | Neutral

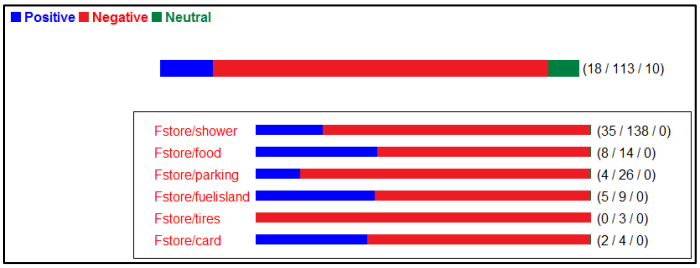
	Type	Body	Weight
1	CONCEPT_RULE	(SENT,"_c{not clean}","_def{Fstoreshower}")	0.08
2	CONCEPT_RULE	(SENT,"_c{dirty}","_def{Fstoreshower}")	0.07
3	CONCEPT_RULE	(SENT,"_c{hair}","_def{Fstoreshower}")	0.09
4	CONCEPT_RULE	(SENT,"_c{worst}","_def{Fstoreshower}")	0.09
5	CONCEPT_RULE	(SENT,"_c{no hot}","_def{Fstoreshower}")	0.04
6	CONCEPT_RULE	(SENT,"_c{cold}","_def{Fstoreshower}")	0.05
7	CONCEPT_RULE	(SENT,"_c{bad}","_def{Fstoreshower}")	0.06
8	CONCEPT_RULE	(SENT,"_c{break@}","_def{Fstoreshower}")	0.06
9	CONCEPT_RULE	(SENT,"_c{_def{TonalKeywordNegative}}","_def{Fstoreshower}")	1

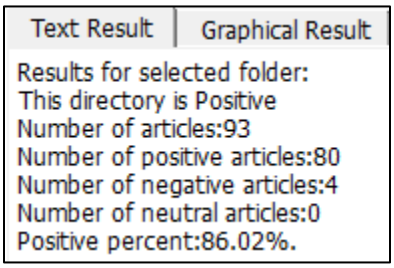
- ✓ Select **Build ⇨ Build Rule-based Model**.
- ✓ Test the folders in the rule-based model and examine the results

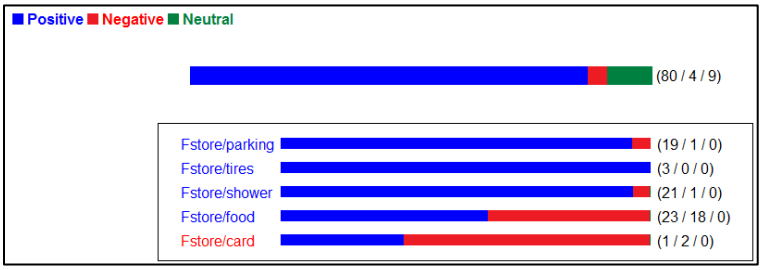
3. **Comment** on the results of the positive and negative tests. **(2 points)**

The rule-based model created better negative precision and results as compared to the statistical rule model but this was not the case for the positive sentiments.



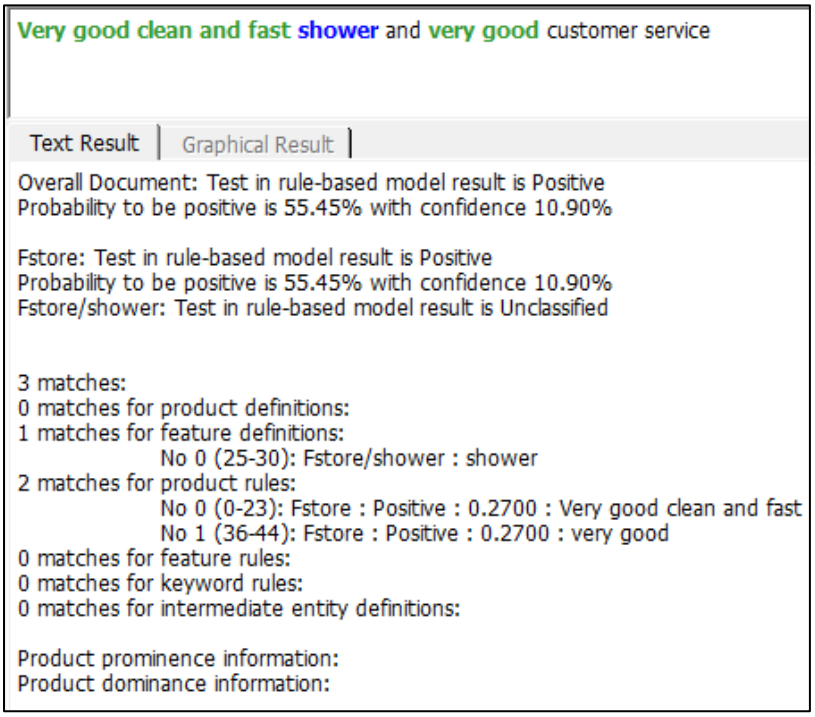






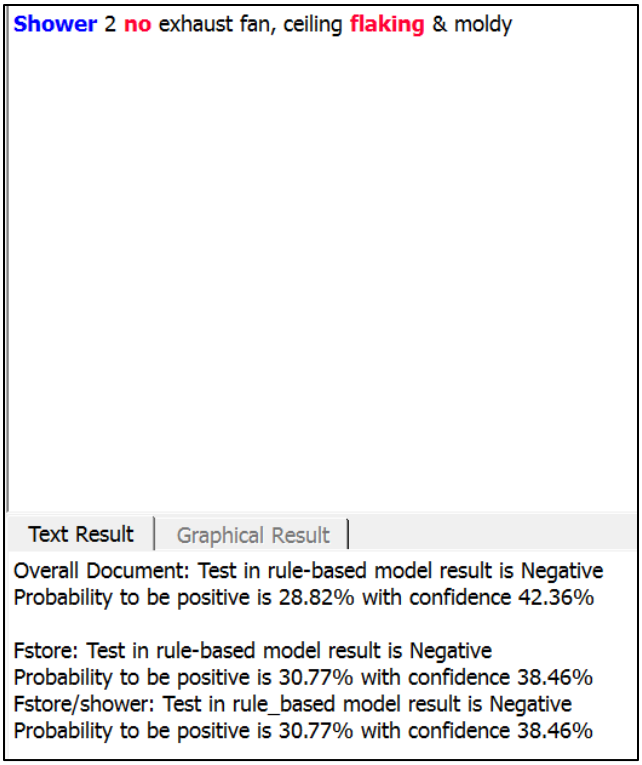
4. Examine how a specific positive comment (within the positive test directory) was tested using the rule-based model. Report screenshot. **(1 point)**

   The rule-based model searches in product and feature definitions, rules, and intermediate definitions for matches. For the below comment we see that there are three matches found – one for feature definition and two for product rules. The green highlighted words have the positive tone for the document.



5. Examine how a specific negative comment was tested (within the negative test directory) using the rule-based model. Report screenshot. **(1 point)**

   Five matches were found for the negative comment below. Two matches for feature definitions, two for feature rules and one for a keyword rule. The feature rules and keyword rules indicate a negative tone. There is 61.9% probability that the document has negative tone. The red highlighted word is the negative word used in connection with the shower.



Deliverables:

As you complete the exercise, create a short report in Microsoft Word and in this report answer the questions in the exercise description. Make sure you print your name, section number, student ID# on the report (front page as well as a running header/footer) and turn-in the report as communicated by your instructor.