```
!pip install turicreate
```

```
from google.colab import drive
drive.mount('/content/gdrive')
```

```
Mounted at /content/gdrive
```

```
import turicreate
```

```
products = turicreate.SFrame('/content/gdrive/My Drive/Turicreate/Week 3/amazon_baby.sframe')
products
```

| name | review | rating |
|---|---|---|
| Planetwise Flannel Wipes | These flannel wipes are OK, but in my opinion ... | 3.0 |
| Planetwise Wipe Pouch | it came early and was not disappointed. i love ... | 5.0 |
| Annas Dream Full Quilt with 2 Shams ... | Very soft and comfortable and warmer than it ... | 5.0 |
| Stop Pacifier Sucking without tears with ... | This is a product well worth the purchase. I ... | 5.0 |
| Stop Pacifier Sucking without tears with ... | All of my kids have cried non-stop when I tried to ... | 5.0 |
| Stop Pacifier Sucking without tears with ... | When the Binky Fairy came to our house, we didn't ... | 5.0 |
| A Tale of Baby's Days with Peter Rabbit ... | Lovely book, it's bound tightly so you may no ... | 4.0 |
| Baby Tracker® - Daily Childcare Journal, ... | Perfect for new parents. We were able to keep ... | 5.0 |
| Baby Tracker® - Daily Childcare Journal, ... | A friend of mine pinned this product on Pinte ... | 5.0 |
| Baby Tracker® - Daily Childcare Journal, ... | This has been an easy way for my nanny to record ... | 4.0 |

[183531 rows x 3 columns]
Note: Only the head of the SFrame is printed.
You can use print_rows(num_rows=m, num_columns=n) to print more rows and columns.

## 1. Build the word count vector for each vector

```
products['word_count'] = turicreate.text_analytics.count_words(products['review'])
products.head()
```

셀 삭제를 실행취소하려면 Ctrl+M Z 또는 수정 메뉴의 실행취소 옵션을 사용하세요.

| name | review | rating | word_count |
|---|---|---|---|
| Planetwise Flannel Wipes | These flannel wipes are OK, but in my opinion ... | 3.0 | {'handles': 1.0, 'stripping': 1.0, ... |
| Planetwise Wipe Pouch | it came early and was not disappointed. i love ... | 5.0 | {'recommend': 1.0, 'highly': 1.0, ... |
| Annas Dream Full Quilt with 2 Shams ... | Very soft and comfortable and warmer than it ... | 5.0 | {'quilt': 1.0, 'of': 1.0, 'the': 1.0, 'than': 1.0, ... |
| Stop Pacifier Sucking without tears with ... | This is a product well worth the purchase. I ... | 5.0 | {'tool': 1.0, 'clever': 1.0, 'approach': 2.0, ... |
| Stop Pacifier Sucking without tears with ... | All of my kids have cried non-stop when I tried to ... | 5.0 | {'rock': 1.0, 'many': 1.0, 'headaches': 1.0, ... |
| Stop Pacifier Sucking without tears with | When the Binky Fairy came | 5.0 | {'thumb': 1.0, 'or': 1.0, 'break': 1.0, 'trying': |

```
selected_words = ['awesome', 'great', 'fantastic', 'amazing', 'love',
                  'horrible', 'bad', 'terrible', 'awful', 'wow', 'hate']
```

| | | | |
|---|---|---|---|
| Baby Tracker® | Perfect for new parents | 5.0 | {'right': 1.0, 'because': |

```
def only_care_select_words(words_source):
    for item in words_source:
        if item not in select_words:
            del words_source[item]

def awesome_count(word_counts):
    return 0 if 'awesome' not in word_counts else word_counts['awesome']


products['awesome'] = products['word_count'].apply(awesome_count)


def word_count_func(word):
    def word_count(word_counts):
        return 0 if word not in word_counts else word_counts[word]
    return word_count

for keyword in selected_words:
    products[keyword]=products['word_count'].apply(word_count_func(keyword))


for keyword in selected_words:
    print('%s: %d' % (keyword, products[keyword].sum()))
```

```
awesome: 4075
great: 59536
fantastic: 1765
amazing: 2726
love: 43867
horrible: 1245
```



```
wow: 461
hate: 1285
```

## 2. Create a new sentiment analysis model using only the selected_words as

```
products = products[products['rating']!= 3]
products['sentiment'] = products['rating'] >= 4


train_data, test_data = products.random_split(0.8, seed=0)


selected_words_model = turicreate.logistic_classifier.create(train_data,target='sentiment',
                                                             features=selected_words,
                                                             validation_set=test_data)
```

```
Logistic regression:
--------------------------------------------------------
Number of examples          : 133448
Number of classes           : 2
Number of feature columns   : 11
Number of unpacked features : 11
Number of coefficients      : 12
Starting Newton Method
--------------------------------------------------------
+-----------+----------+--------------+--------------------+----------------------+
| Iteration | Passes   | Elapsed Time | Training Accuracy | Validation Accuracy |
+-----------+----------+--------------+--------------------+----------------------+
| 1         | 2        | 1.222809     | 0.847401          | 0.845874            |
| 2         | 3        | 1.382795     | 0.847514          | 0.846085            |
| 3         | 4        | 1.543680     | 0.847626          | 0.846115            |
| 4         | 5        | 1.704688     | 0.847708          | 0.846385            |
| 5         | 6        | 1.864652     | 0.847708          | 0.846385            |
| 6         | 7        | 2.031830     | 0.847708          | 0.846385            |
+-----------+----------+--------------+--------------------+----------------------+
SUCCESS: Optimal solution found.
```

```
swm_sorted = selected_words_model.coefficients.sort('value', ascending=False)
swm_sorted.print_rows(num_rows=12)
```

```
+-------------+-------+-------+---------------------+----------------------+
|     name    | index | class |        value        |        stderr        |
+-------------+-------+-------+---------------------+----------------------+
|     love    |  None |   1   |  1.3592688669224757 | 0.028068300152099997 |
| (intercept) |  None |   1   |  1.3365913848878148 | 0.008929969787657532 |
|   awesome   |  None |   1   |  1.1335346660341103 |  0.08399643983187567 |
|   amazing   |  None |   1   |  1.1000933113659914 |  0.09954776260465965 |
|  fantastic  |  None |   1   |  0.8858047568813963 |  0.11167591293399713 |
|    great    |  None |   1   |  0.8630655001195999 | 0.018955052444380473 |
|     wow     |  None |   1   | -0.0095382360677122 |  0.1604641122471162  |
|     bad     |  None |   1   | -0.9914778800650894 | 0.03848428664699065  |
|     hate    |  None |   1   | -1.3484407222463402 | 0.07715698604297318  |
+-------------+-------+-------+---------------------+----------------------+
[12 rows x 5 columns]
```

셀 삭제를 실행취소하려면 Ctrl+M Z 또는 수정 메뉴의 실행취소 옵션을 사용하세요.

## 3. Comparing the accuracy of different sentiment analysis model

```
selected_words_model.evaluate(test_data)
```

```
{'accuracy': 0.8463848186404036,
 'auc': 0.6935096220934976,
 'confusion_matrix': Columns:
        target_label    int
        predicted_label int
        count   int

 Rows: 4

 Data:
 +--------------+-----------------+-------+
 | target_label | predicted_label | count |
 +--------------+-----------------+-------+
 |      1       |        0        |  159  |
 |      0       |        0        |  371  |
 |      0       |        1        |  4957 |
 |      1       |        1        | 27817 |
 +--------------+-----------------+-------+
 [4 rows x 3 columns],
 'f1_score': 0.9157860082304526,
 'log_loss': 0.3962265467087835,
 'precision': 0.8487520595594068,
 'recall': 0.9943165570488991,
 'roc_curve': Columns:
        threshold       float
        fpr     float
        tpr     float
        p       int
        n       int

 Rows: 1001

 Data:
 +-----------+--------------------+-----+-------+------+
 | threshold |        fpr         | tpr |   p   |  n   |
 +-----------+--------------------+-----+-------+------+
 |    0.0    |        1.0         | 1.0 | 27976 | 5328 |
 |   0.001   | 0.9994369369369369 | 1.0 | 27976 | 5328 |
 |   0.002   | 0.9992492492492493 | 1.0 | 27976 | 5328 |
 |   0.003   | 0.9990615615615616 | 1.0 | 27976 | 5328 |
 |   0.004   | 0.9990615615615616 | 1.0 | 27976 | 5328 |
 |   0.005   | 0.9981231231231231 | 1.0 | 27976 | 5328 |
 |   0.006   | 0.9981231231231231 | 1.0 | 27976 | 5328 |
 |   0.007   | 0.9979354354354354 | 1.0 | 27976 | 5328 |
 |   0.008   | 0.9973723723723724 | 1.0 | 27976 | 5328 |
 |   0.009   | 0.9971846846846847 | 1.0 | 27976 | 5328 |
 +-----------+--------------------+-----+-------+------+
 [1001 rows x 5 columns]
```

셀 삭제를 실행취소하려면 Ctrl+M Z 또는 수정 메뉴의 실행취소 옵션을 사용하세요.

columns.}

## 4. Interpreting the difference in performance between the models

```
diaper_champ_reviews = products[products['name'] == 'Baby Trend Diaper Champ']
diaper_champ_reviews.head()
```

| name | review | rating | word_count | awesome | great | fantastic |
|---|---|---|---|---|---|---|
| Baby Trend Diaper Champ | Ok - newsflash. Diapers are just smelly. We've ... | 4.0 | {'convenient': 1.0, 'more': 1.0, 'trash': ... | 0.0 | 0.0 | 0.0 |
| Baby Trend Diaper Champ | This is a good product to start and very easy to ... | 3.0 | {'bad': 1.0, 'smells': 1.0, 'because': 1.0, ... | 0.0 | 0.0 | 0.0 |
| Baby Trend Diaper Champ | My husband and I selected the Diaper "Champ" ma ... | 1.0 | {'system': 1.0, 'try': 1.0, 're': 1.0, 'still': ... | 0.0 | 0.0 | 0.0 |
| Baby Trend Diaper Champ | Excellent diaper disposal unit. I used it in ... | 5.0 | {'nose': 1.0, 'for': 2.0, 'investment': 1.0, ... | 0.0 | 0.0 | 0.0 |
| Baby Trend Diaper Champ | We love our diaper champ. It is very easy to use ... | 5.0 | {'out': 1.0, 'pull': 1.0, 'open': 1.0, 'pail': ... | 0.0 | 0.0 | 0.0 |
| Baby Trend Diaper Champ | Two girlfriends and two family members put me ... | 5.0 | {'winter': 1.0, 'outside': 1.0, 'day': ... | 0.0 | 0.0 | 0.0 |
| Baby Trend Diaper Champ | I waited to review this until I saw how it ... | 4.0 | {'mom': 1.0, 'my': 1.0, 'empty': 2.0, 'poop': ... | 0.0 | 0.0 | 0.0 |
| Baby Trend Diaper Champ | I have had a diaper genie for almost 4 years since ... | 1.0 | {'yuck': 1.0, 'clean': 1.0, 'all': 1.0, 'tra ... | 0.0 | 0.0 | 0.0 |
| Baby | I originally | 5.0 | {'price': 1.0, | 0.0 | 0.0 | 0.0 |

셀 삭제를 실행취소하려면 Ctrl+M Z 또는 수정 메뉴의 실행취소 옵션을 사용하세요.

r_champ_reviews, ou
g=False)

```
diaper_champ_reviews[diaper_champ_reviews['rating']==5]
```

| name | review | rating | word_count | awesome | great | fantastic |
|---|---|---|---|---|---|---|
| Baby Trend Diaper Champ | Love it, love it, love it! This lives up to ... | 5.0 | {'parents': 1.0, 'new': 1.0, 'recommended': 1.0, ... | 0.0 | 0.0 | 0.0 |
| Baby Trend Diaper Champ | I received my Diaper Champ at my baby shower ... | 5.0 | {'second': 2.0, 'have': 1.0, 'are': 1.0, 'pull': ... | 0.0 | 0.0 | 0.0 |
| Baby Trend Diaper Champ | Let me just say, I LOVE THIS PRODUCT!! I used ... | 5.0 | {'friends': 1.0, 'feces': 1.0, 'tank': 1.0, ... | 0.0 | 1.0 | 1.0 |
| Baby Trend Diaper Champ | I love this diaper pale and wouldn't dream of ... | 5.0 | {'easy': 1.0, 'how': 1.0, 'like': 1.0, 'on': 1.0, ... | 0.0 | 2.0 | 0.0 |
| Baby Trend Diaper Champ | Works great - no smells. LOVE that it uses reg ... | 5.0 | {'plastic': 1.0, 'in': 1.0, 'each': 1.0, ... | 0.0 | 2.0 | 0.0 |
| Baby Trend Diaper Champ | I have been using this diaper pail for 41/2 ... | 5.0 | {'recommend': 1.0, 'would': 1.0, 'with': ... | 0.0 | 2.0 | 0.0 |
| Baby Trend Diaper Champ | This is absolutely, by far, the best diaper ... | 5.0 | {'praises': 1.0, 'sing': 1.0, 'glad': 1.0, ... | 0.0 | 0.0 | 0.0 |
| Baby | I've | 5.0 | {'diapers': 1.0 | 0.0 | 0.0 | 0.0 |

```
selected_words_model.predict(diaper_champ_reviews[0:1], output_type='probability')
```

```
dtype: float
Rows: 1
[0.998125362333512]
```

```
items = diaper_champ_reviews[1]['word_count']
for item in sorted(items, key=items.get, reverse=True):
```

셀 삭제를 실행취소하려면 Ctrl+M Z 또는 수정 메뉴의 실행취소 옵션을 사용하세요.

```
the 3.0
love 3.0
this 3.0
to 3.0
your 2.0
i 2.0
can 2.0
```

```
use 2.0
if 2.0
parents 1.0
new 1.0
recommended 1.0
no 1.0
then 1.0
problem 1.0
fingertips 1.0
finger 1.0
side 1.0
just 1.0
that 1.0
out 1.0
clean 1.0
or 1.0
instead 1.0
course 1.0
diapers 1.0
house 1.0
smell 1.0
don 1.0
up 1.0
lives 1.0
s 1.0
we 1.0
t 1.0
diaper 1.0
change 1.0
but 1.0
found 1.0
unless 1.0
also 1.0
all 1.0
any 1.0
already 1.0
purchased 1.0
hard 1.0
bags 1.0
is 1.0
trash 1.0
around 1.0
cannot 1.0
hype 1.0
which 1.0
why 1.0
open 1.0
particular 1.0
pail 1.0
be 1.0
long 1.0
```

✓ 0초 오후 1:53에 완료됨

셀 삭제를 실행취소하려면 Ctrl+M Z 또는 수정 메뉴의 실행취소 옵션을 사용하세요.