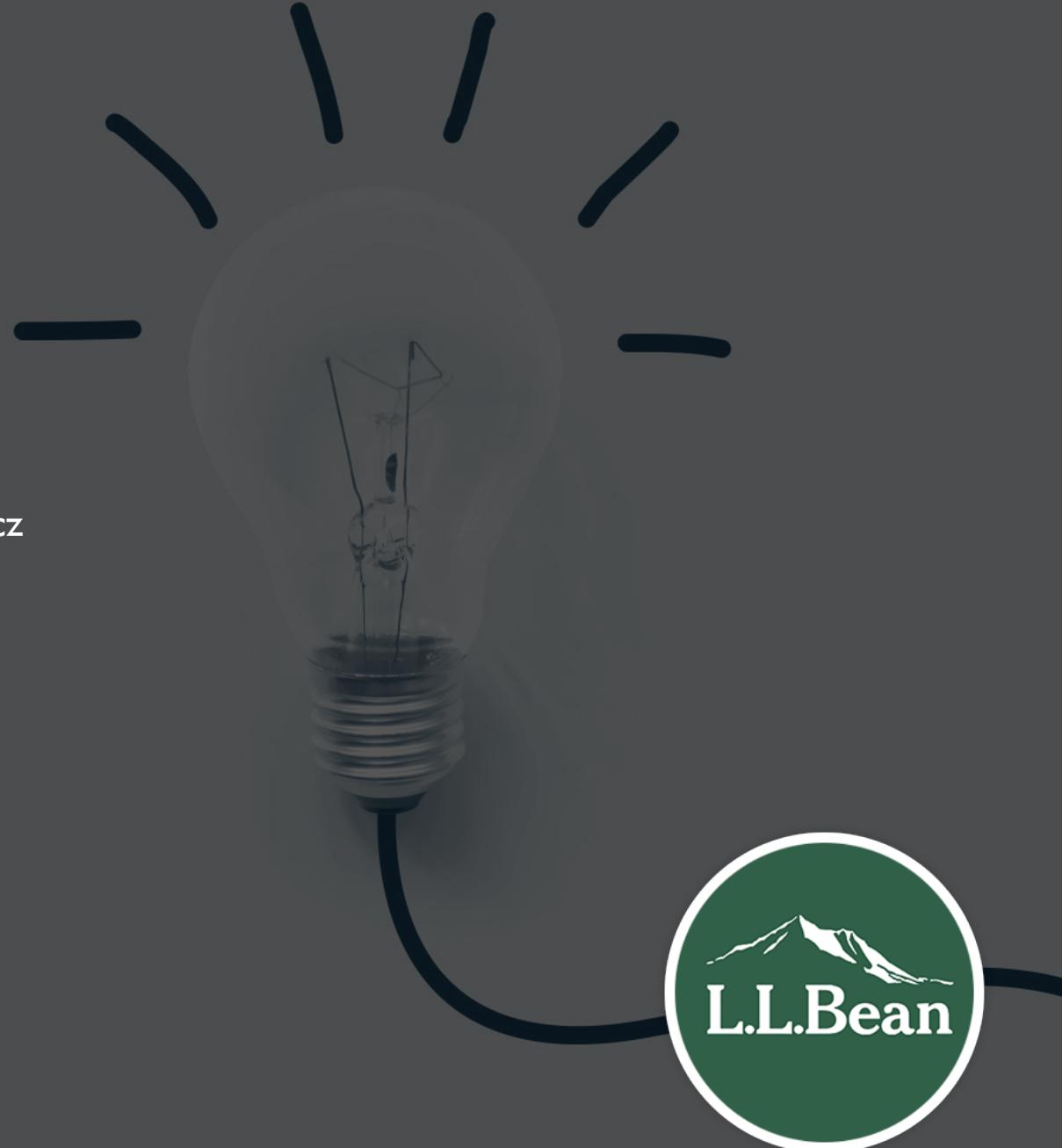


DERIVING INSIGHTS FROM LL BEAN CUSTOMER REVIEW DATA

Ryan Goebel
Tim Przybylowicz

ALY6040



A close-up photograph of a person's hands writing in a spiral-bound notebook with a blue pen. The person is wearing a dark long-sleeved shirt and a black smartwatch on their left wrist. The background is slightly blurred, showing what appears to be a whiteboard or a wall.

AGENDA

01

LL BEAN OVERVIEW & OBJECTIVES

02

DATA REQUIREMENTS

03

DATA COLLECTION & PRE-PROCESSING

04

SENTIMENT CLASSIFICATION

05

NATURAL LANGUAGE PROCESSING

LL BEAN

Industry:

Retail Company

Products:

Outdoor Apparel, Equipment, and Footwear

Awards:

Voted 4th in the United States in Company Reputation (2019, Axios Harris)

Goals:

Expand their Footprint and Product Offerings



BUSINESS OBJECTIVES*

LL BEAN



GROW THE BRAND



MAINTAIN REPUTATION



UNDERSTAND THEIR CUSTOMER



*These in no way reflect the actual opinions of LL Bean

FEATURED REVIEW



Excellent T

Size Range: Regular

Size: X-Large

Fit: True to Size

My husband is very picky about his T Shirts, and this one has him rejecting all others. The fit is perfect, the fabric is excellent, and it really does not shrink and holds its shape. Go as fast as you can and order this shirt in every color! I have and gotten rid of everything else.

[Post a Comment](#)



[Show Comments \(3\)](#)

[Report Inappropriate Content](#)

November 28, 2018

Ritab

Jacksonville, FL

Featured

✓ **Verified Purchase**

74 of 79 found this review helpful.

Was this helpful? [Yes](#) | [No](#)

SOLUTION REQUIREMENTS

For this project, we focused on helping LL Bean....

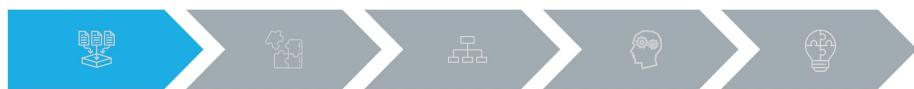
...grow the Brand, while maintaining reputation and customer satisfaction.

By Answering Key Questions like:

- How can we quantify customer satisfaction? **Sentiment Analysis**
- Can we identify key product deficiencies before they risk LL Bean's reputation? **Natural Language Processing**

To support these types of analysis, we required the following data:

- Product Description
- Customer Review Rating
- Customer Review Text



DATA COLLECTION – SCRAPING OVERVIEW

Shop All Clothing

Sort by: Recommended

1413 Items

Women's L.L.Bean Sweater Fleece Pullover \$64.99-\$89.95 ★★★★☆

Women's Vista Camp Pants \$49.99-\$74.95 ★★★★☆

Women's Vista Camp Pants, Cropped \$54.99-\$74.95 ★★★★☆

Women's Perfect Fit Pants, Straight Leg \$29.99-\$44.95 ★★★★☆

Women's Perfect Fit Pants, Cropped \$39.95-\$44.95 ★★★★☆

Women's L.L.Bean Sweater Fleece Full-Zip Jacket \$89.99 ★★★★☆

Women's Comfort Trail Shorts \$54.95-\$59.95 ★★★★☆

Women's Prima Cotton Tee, Short-Sleeve Crewneck \$22.95-\$27.95 ★★★★☆

Women's Prima Cotton Tee, Long-Sleeve Crewneck \$26.95-\$32.95 ★★★★☆

Women's Summer Knit Dress, Short-Sleeve Print \$59.95-\$64.95 ★★★★☆

Women's Trail Model Rain Jacket \$59-\$109 ★★★★☆

Women's Mountain Classic Anorak, Multi-Color \$59-\$69 ★★★★☆

Open Link

FEATURED REVIEW

COOOOOOOZY!

January 16, 2020

TREx56 Loveland, OH Featured ✓ Verified Purchase

Size Range: Regular
Size: Large
Fit: Somewhat Large
Customer is 5'8" and 135 lbs.

Super warm and comfortable! My wife wears it constantly. Great around the house, breathable and soft! Love the warm raspberry color!

Post a Comment + Show Comment (1) Report Inappropriate Content

25 of 28 found this review helpful.
Was this helpful? Yes | No

Identify Tags & Scrape Information

Size Range: Regular
Size: Large
Fit: Somewhat Large

`div.ReviewBody_review-text` 784 x 53

Super warm and comfortable! My wife wears it constantly. Great around the house, breathable and soft! Love the warm raspberry color!

Post a Comment + Show Comment (1) Report Inappropriate Content

Open Next Link

DATA COLLECTION – FEATURE EXTRACTION

Key Limitation: Scraping took >24 hours total

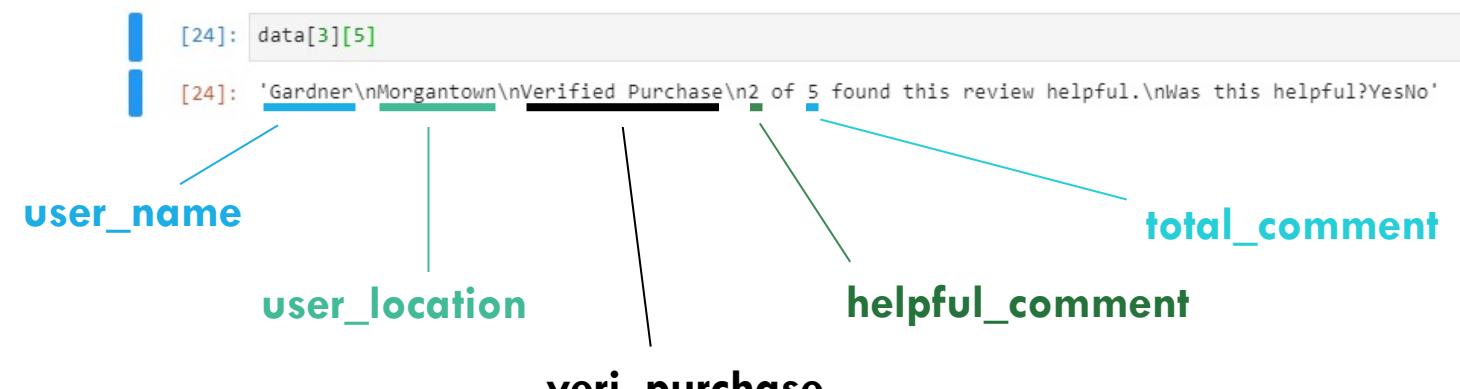
Sacrifices: Post-processing required to pull features from raw text.

review_rating

```
[19]: data[6]
[19]: 0      Rating: 3 out of 5 stars
      1      Rating: 1 out of 5 stars
      2      Rating: 3 out of 5 stars
      3      Rating: 5 out of 5 stars
      4      Rating: 2 out of 5 stars
      ...
8539   Rating: 5 out of 5 stars
8540   Rating: 5 out of 5 stars
8541   Rating: 5 out of 5 stars
8542   Rating: 5 out of 5 stars
8543   Rating: 5 out of 5 stars
```

[24]: data[3][5]

[24]: 'Gardner\nMorgantown\nVerified Purchase\n2 of 5 found this review helpful.\nWas this helpful?YesNo'



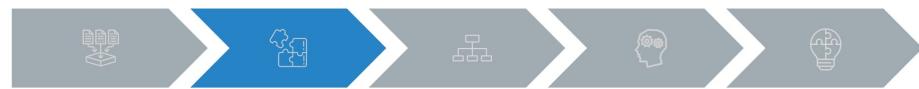
user_name

user_location

veri_purchase

helpful_comment

total_comment



CLEANED DATASET OVERVIEW

Number of Reviews: 205,373

Number of Products: 837

Highest Rated Product*: Men's Wicked Good Flannel Shirt, Traditional Fit, Houndstooth
(Avg Rating: 4.73)

Lowest Rated Product*: Women's Wrinkle-Free Bayside Pants, Favorite Fit
(Avg Rating: 3.50)



PROBLEM – RATING AMBIGUITY

LL Bean Employs a 5-Star Rating System

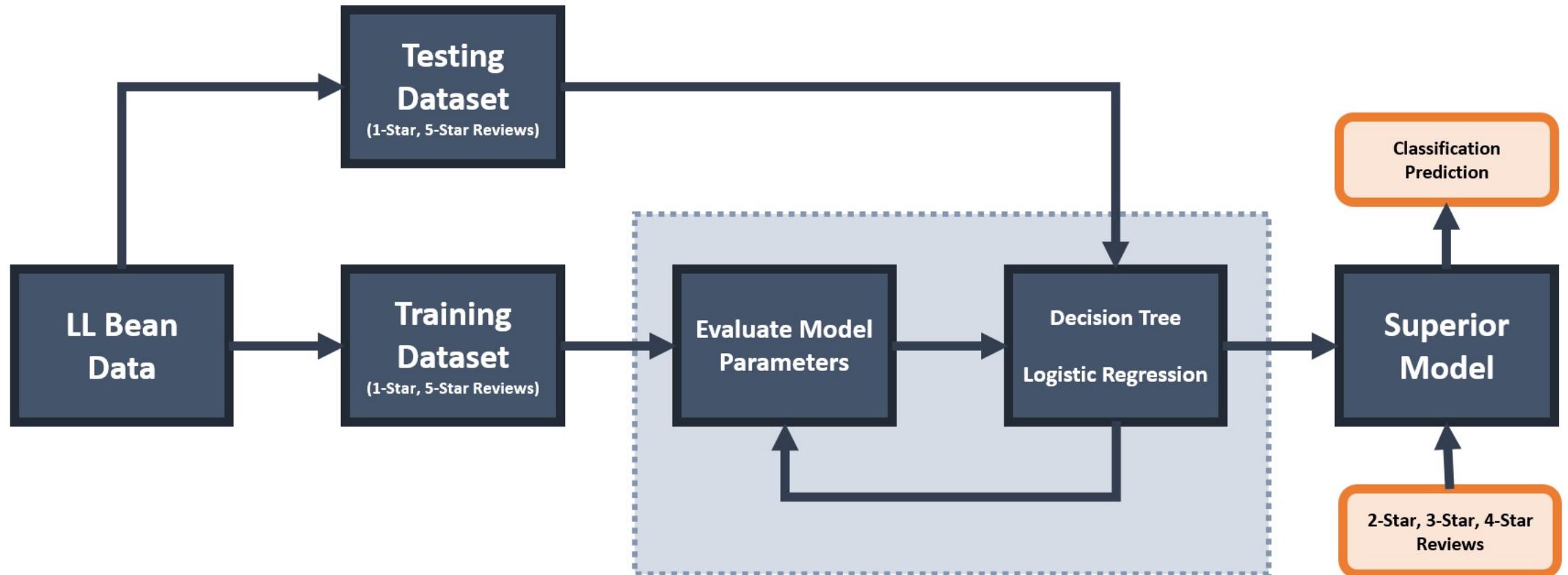
- One Star = Would NOT Recommend
- Five Star = Would Recommend
- Two, Three, Four Star = ?

Solution:

- Remove Ambiguity.
- **Classify each review into Positive or Negative based on the review title and text.**



CLASSIFICATION MODEL - METHODOLOGY



CLASSIFICATION MODEL - RESULTS

```
[1] =====  
[1] "Logistic Regression Classifier Results"  
      Predicted  
Observed   negative  positive  
negative      2530     1544  
positive       381    27464  
[1] "The accuracy of the Regression Model was 93.97 %"  
[1] "The sensitivity of the Regression Model was 98.63 %"  
[1] =====
```

```
[1] =====  
[1] "Pruned Decision Tree Classifier Results"  
      Predicted  
observed   negative  positive  
negative      1662     2412  
positive       561    27284  
[1] "The accuracy of the Decision Tree Model was 90.69 %"  
[1] "The sensitivity of the Regression Model was 97.99 %"  
[1] =====
```

User Rating = 4, Our Classification = Negative

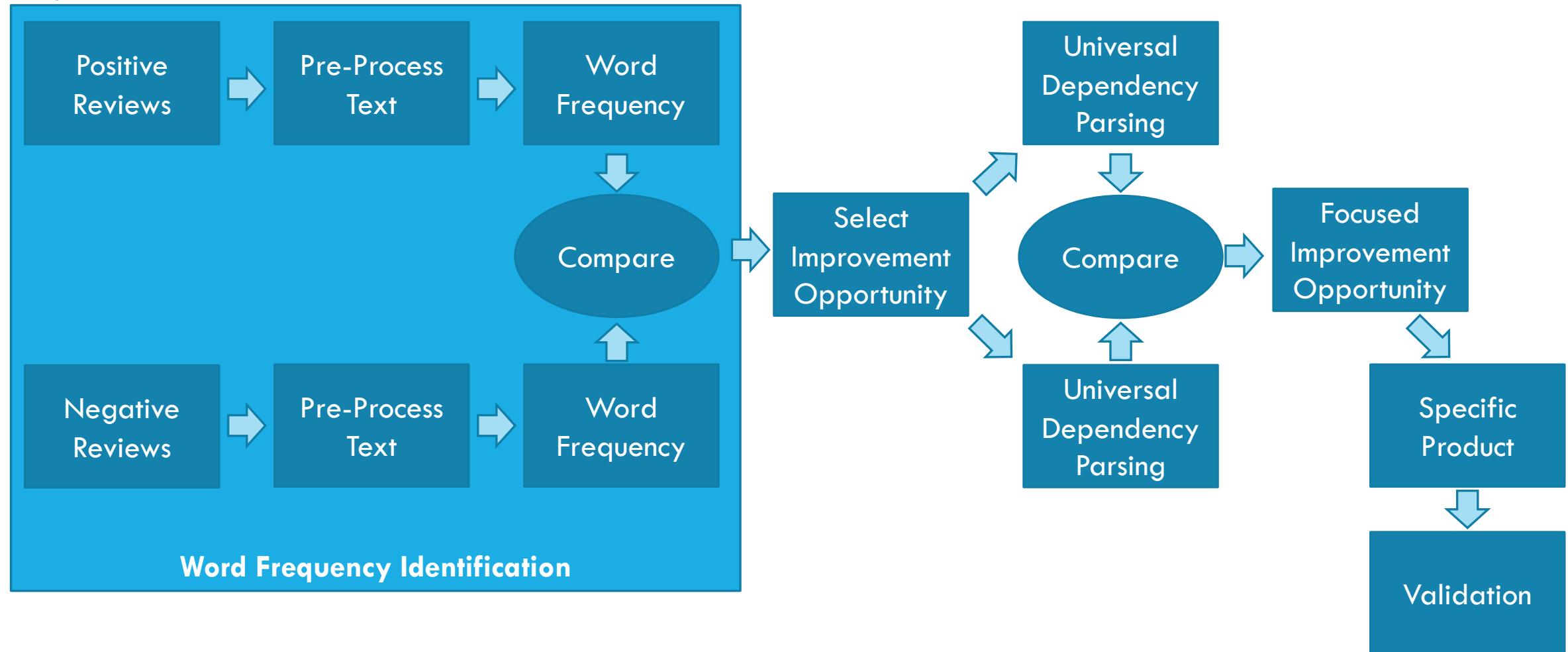
```
> LLB_4$reviewText[310]  
[1] "Basic tee shirt to wear for garden work. Fabric is thinner than shirts I purchased before, so these will be t  
he last I buy from LLBean. I can get the same thin shirts at local stores. My old ones held on for several season  
s; this one may not make more than one. Recommend with reservations."
```

```
> LLB_4$reviewText[506]  
[1] "Washing in cold water and drying medium heat takes sizing and changes shape of garment. very disappointing"
```

```
> LLB_4$reviewText[732]  
[1] "At 65, I don't need to show off my buns. So I would appreciate a little more room in this area. I have kept t  
he pants because they appear to be durable and I can put on a long shirt to cover my buns. One more thing: the pan  
ts make you look slender and tall. They are comfortable in spite of the tightness on the buns, the material stretc  
h enough to offer some room."
```



NATURAL LANGUAGE PROCESSING - OVERVIEW



MOST FREQUENT WORDS – TEXT MINING



Create separate corpora for positive and negative reviews

Pre-process text:

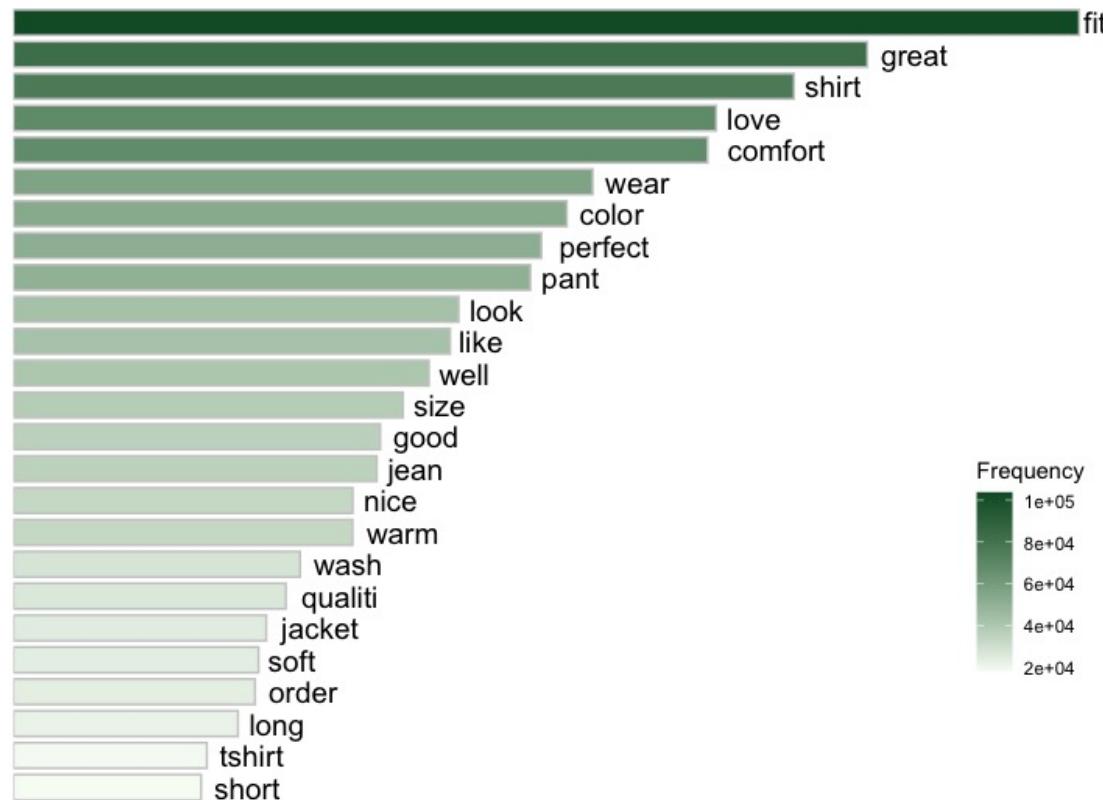
- Lower case
- Remove numbers, stop words, and punctuation
- Combine words such as "t shirt", "tee shirt", and "teeshirt" to "tshirt"
- Stem words and Strip white space

Convert to Document Term Matrix

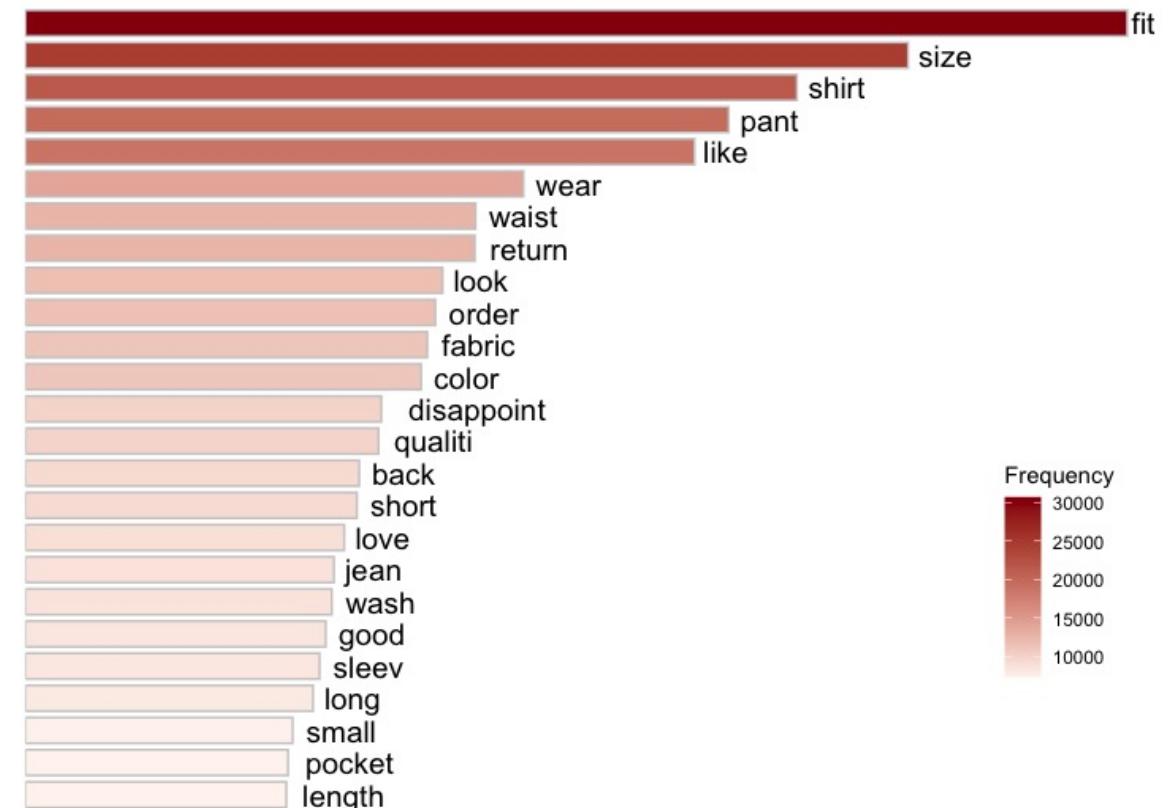
Determine most frequent words in Positive Reviews and in Negative Reviews

POS & NEG REVIEWS - MOST FREQUENT WORDS

Top 25 Words Used in Positive LL Bean Reviews

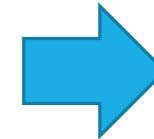


Top 25 Words Used in Negative LL Bean Reviews



WORD FREQUENCY COMPARISON

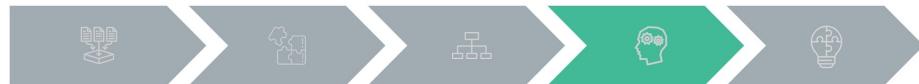
	Positive Reviews	Negative Reviews	Rel. Freq. Diff.
Most Positive			
great	83106	6429	0.0175
comfort	67562	6528	0.0133
love	68393	8878	0.0118
perfect	51376	4948	0.0101
well	40433	5740	0.0066
color	53860	11019	0.0063
warm	33001	3832	0.0060
fit	103729	30687	0.0051
Most Negative			
return	3525	12514	-0.0083
size	37905	24584	-0.0080
disappoint	2344	9912	-0.0067
waist	16239	12534	-0.0049
back	8828	9295	-0.0045
fabric	17311	11192	-0.0036
way	5756	6647	-0.0034
small	8874	7442	-0.0031



Positive Reviews

right
easi
look
shirt
Warm
weight
husband
light
weather
casual
jacket
good
perfect
well
sweater
comfi
layer
excel
jean
soft
winter
flannel
true
enough
day
beauti
love
jean
color
thank
great
nice
favorit
wear
sleeve
leg
thin
poor
problem
arm
tric
cut
hip
order
use
return
tight
old
back
way
however
waist
inch
thought
disappoint
small
chang
fabric
materi
like
seem
button
short
big
much
zipper

Negative Reviews

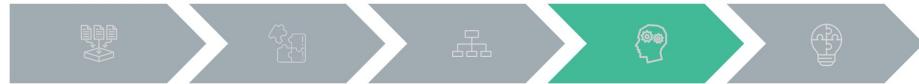


WORD FREQUENCY COMPARISON

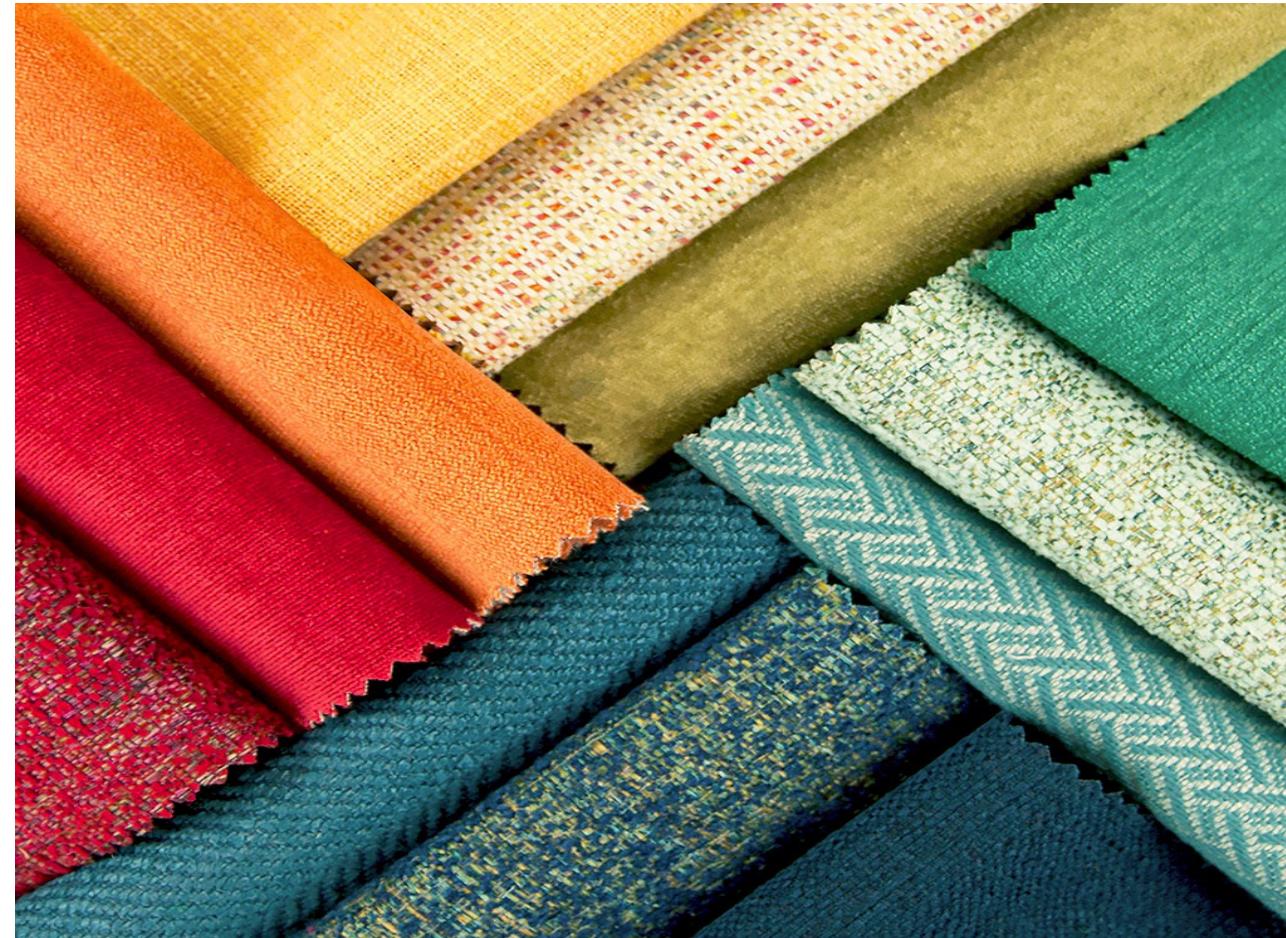
Investigate:

- **"Fabric"** and **"Material"** have higher relative frequencies in Negative Reviews
 - 15.2% of Positive Reviews
 - 30.2% of Negative Reviews





INVESTIGATE NEGATIVE REVIEWS MENTIONING “FABRIC” AND/OR “MATERIAL”



UNIVERSAL DEPENDENCY PARSING

UDPipe package in R

- Tokenization
- POS tagging
- Lemmatization
- Dependency Parsing using universal dependency treebank models

Find nominal subjects & associated adjectives

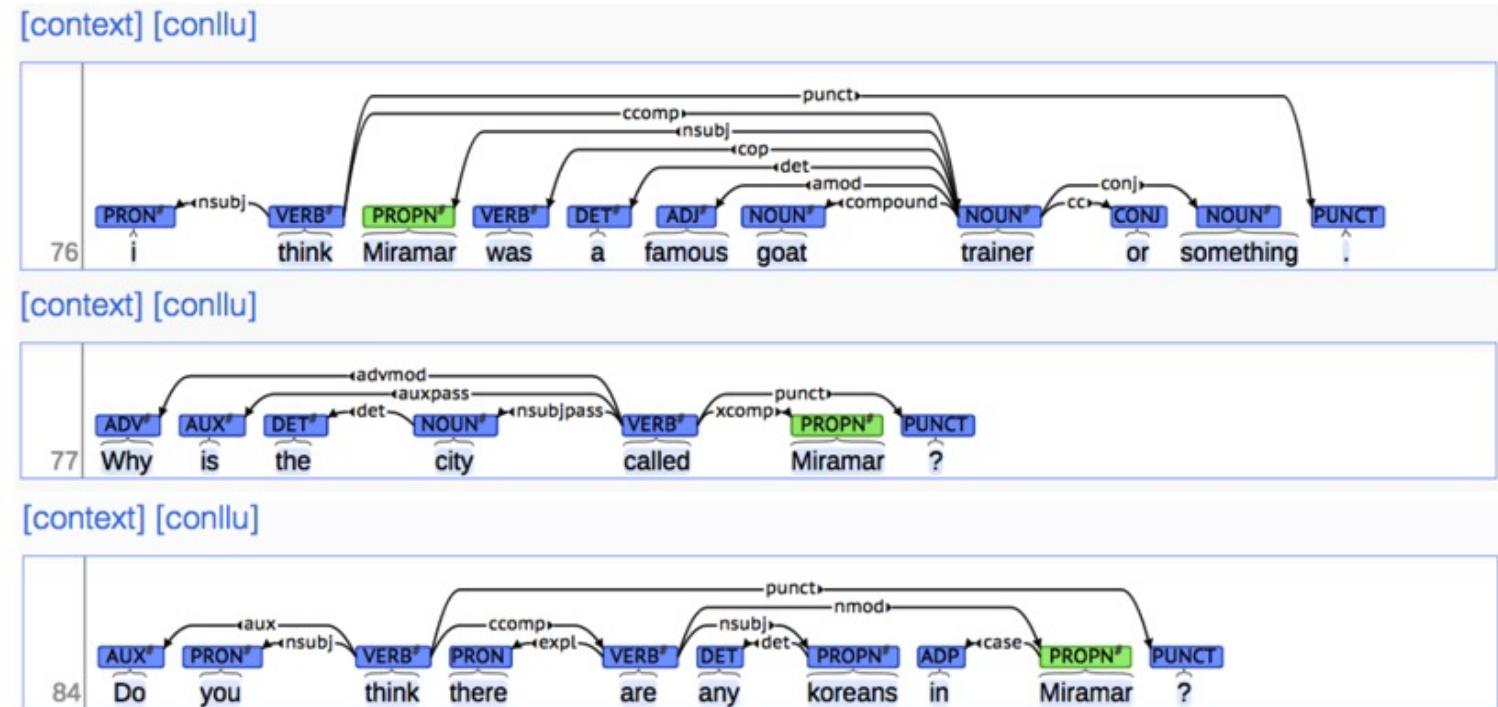


Diagram from: <https://www.cellstrat.com/2020/04/16/dependency-parsing-for-nlp/>

UDP: REVIEWS WITH “FABRIC” AND/OR “MATERIAL”

```
> head(posneg[order(-posneg$Difference),],10)
```

	Phrase	Positive Reviews	Negative Reviews	Difference
6415	soft fabric	1999	309	1690
5110	perfect fit	590	46	544
5140	perfect length	314	53	261
1218	comfortable fabric	299	71	228
2989	great fit	281	69	212
6712	substantial fabric	216	58	158
7248	true fit	162	24	138
1263	comfortable pants	164	34	130
2981	great fabric	292	166	126
2200	excellent quality	132	22	110

```
> head(posneg[order(posneg$Difference),],10)
```

	Phrase	Positive Reviews	Negative Reviews	Difference
6968	thin fabric	106	415	-309
7017	thinner fabric	49	191	-142
5991	short sleeve	39	106	-67
7135	tight waist	25	85	-60
2398	fine fit	14	73	-59
6567	stiff fabric	10	68	-58
1699	different fabric	42	92	-50
7833	wide leg	15	65	-50
2395	fine fabric	6	55	-49
7123	tight sleeve	15	63	-48

Positive Reviews

wonderful pants substantial weight
 wonderful pants accurate sizing
 available color comfortable jeans
 happy husband warm fabric reasonable price
 soft cotton durable fabric comfortable tee
 perfect style great color great fabric deep pockets perfect medium
 light fabric perfect shirt vibrant color beautiful fabric
 right weight excellent fabric relaxed fit right length thick fabric comfortable waistband
 soft feel comfortable shirt perfect weight beautiful color
 perfect size rich color excellent quality right fit great pants strong fabric
 perfect jacket comfortable short enough fabric soft shirt substantial fabric
 breathable fabric perfect pants comfortable fabric
 best shirt sturdy fabric true fit perfect length silly fabric
 right price comfortable pants perfect length perfect sleeve
 best fabric outstanding fabric perfect fabric great fit perfect fit
 perfect cut excellent fit cool fabric comfortable fit
 perfect jeans great shirt

true sizing
 soft fabric
 thin fabric
 thinner fabric
 good quality large pants short rise fine length big pants awfully fit
 stiff fabric terrible fit short length big shirt
 tight sleeves nice fabric small waist flimsy fabric wider leg
 long sleeve fine fit awful fabric huge shirt awful fabric
 heavy fabric inferior fabric tight fit good fabric same fabric different fit
 coarse fabric high waist short shirt baggy leg tight pants short arm
 huge waist rough fabric big waist huge leg different pants
 smaller size ok fabric poor quality same quality high rise lighter fabric
 fine quality baggy pants more fabric straight leg large waist
 tight jeans strange fit scratchy fabric thing fabric smaller waist

Negative Reviews

UDP: REVIEWS WITH “FABRIC” AND/OR “MATERIAL”

```
> head(posneg[order(-posneg$Difference),],10)
```

	Phrase	Positive Reviews	Negative Reviews	Difference
6415	soft fabric	1999	309	1690
5110	perfect fit	590	46	544
5140	perfect length	314	53	261
1218	comfortable fabric	299	71	228
2989	great fit	281	69	212
6712	substantial fabric	216	58	158
7248	true fit	162	24	138
1263	comfortable pants	164	34	130
2981	great fabric	292	166	126
2200	excellent quality	132	22	110

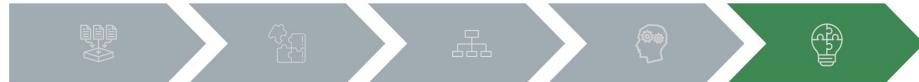
```
> head(posneg[order(posneg$Difference),],10)
```

	Phrase	Positive Reviews	Negative Reviews	Difference
69	thin fabric	106	415	-309
70	thinner fabric	49	191	-142
5951	short sleeve	39	106	-67
7135	tight waist	25	85	-60
2398	fine fit	14	73	-59
6567	stiff fabric	10	68	-58
1699	different fabric	42	92	-50
7833	wide leg	15	65	-50
2395	fine fabric	6	55	-49
7123	tight sleeve	15	63	-48

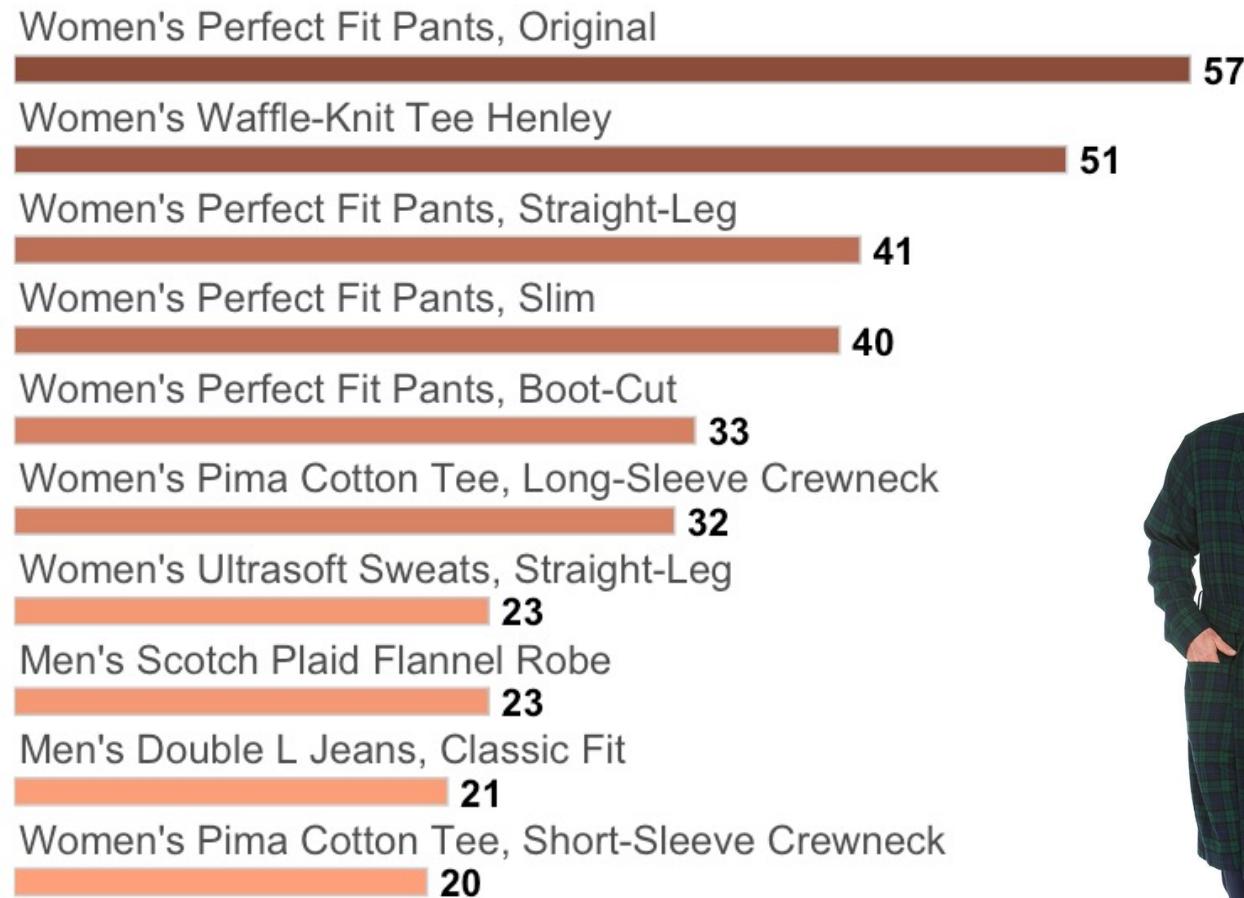
Positive Reviews

wonderful pants substantial weight
 happy husband warm fabric reasonable price
 soft cotton durable fabric great fabric deep pockets comfortable tee
 perfect style great color perfect color great length perfect medium
 light fabric perfect shirt vibrant color beautiful fabric
 right weight excellent fabric relaxed fit right length thick fabric comfortable waistband
 soft feel comfortable shirt perfect weight beautiful color
 perfect size rich color excellent quality right fit great pants strong fabric
 perfect jacket comfortable short enough fabric soft shirt substantial fabric
 breathable fabric perfect pants comfortable fabric
 best shirt sturdy fabric true fit perfect length silly fabric
 right price comfortable pants perfect length perfect sleeves
 best fabric outstanding fabric perfect fabric great fit perfect
 perfect cut excellent fit cool fabric comfortable fit
 perfect jeans great shirt
 perfect fit
soft fabric
 thin fabric
 thinner fabric
 good quality
 large pants short rise
 fine length
 big pants
 awful fit
 tight sleeve
 heavy fabric
 inferior fabric
 coarse fabric
 huge waist
 smaller size
 fine quality
 same quality
 baggy pants
 tight jeans
 thing fabric
 wide leg
 stiff fabric
 short length
 big shirt
 ok fit
 small waist
 flimsy fabric
 inferior fabric
 short shirt
 huge shirt
 different fit
 cheap fabric
 baggy leg
 tight pants
 short arm
 huge leg
 odd fit
 fine waist
 large fit
 lighter fabric
 straight leg
 scratchy fabric
 smaller waist
 tight one
 terrible fit
 tight shirt
 big shirt
 wider leg
 awful fabric
 different fit
 tight fit
 short arm
 huge leg
 odd fit
 fine waist
 large fit
 lighter fabric
 straight leg
 scratchy fabric
 smaller waist

Negative Reviews

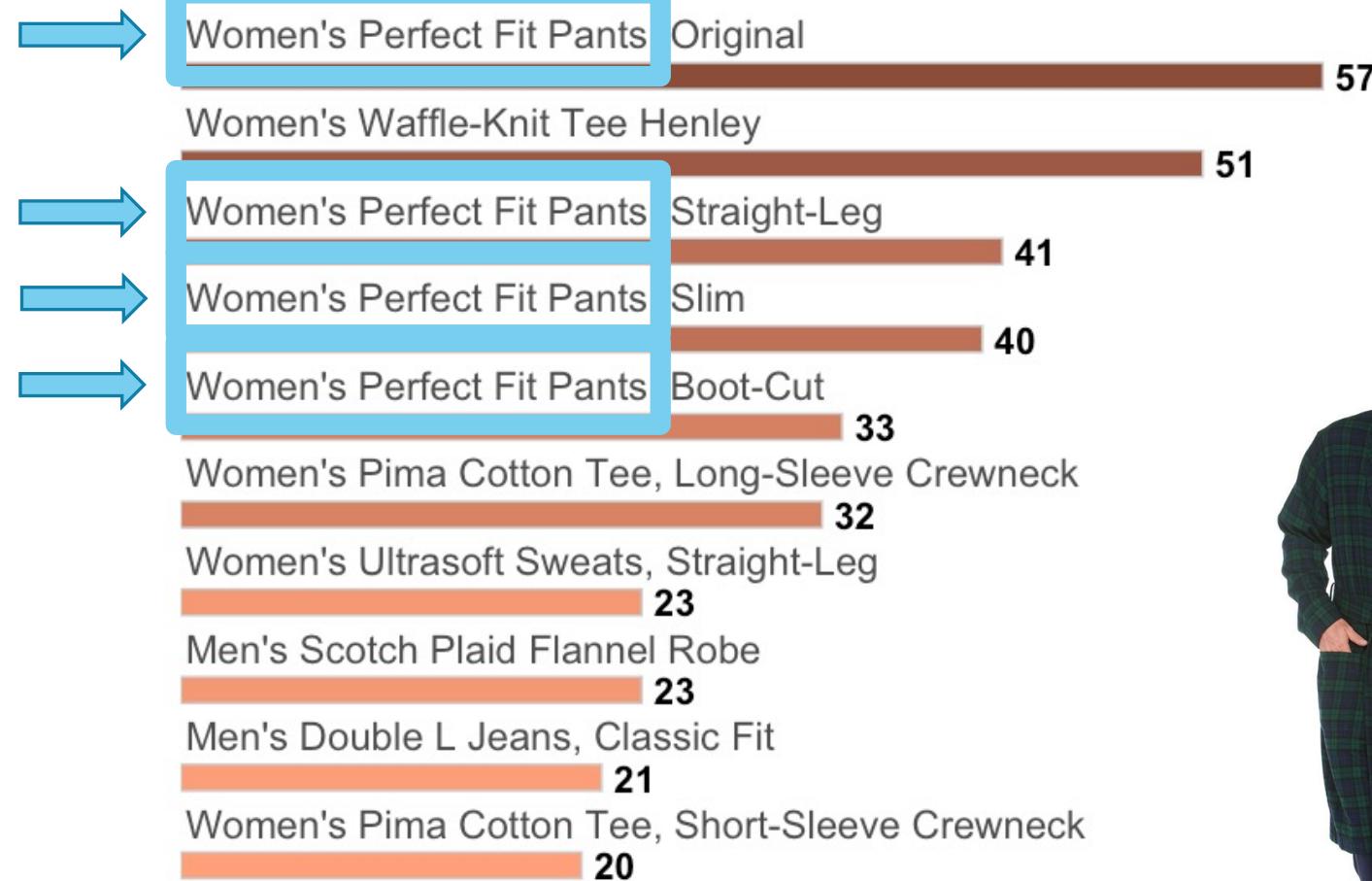


PRODUCTS WITH THE MOST NEGATIVE REVIEWS MENTIONING “THIN(NER) FABRIC” OR “THIN(NER) MATERIAL”



Photos from: llbean.com

PRODUCTS WITH THE MOST NEGATIVE REVIEWS MENTIONING “THIN(NER) FABRIC” OR “THIN(NER) MATERIAL”



Photos from: llbean.com

VALIDATION: SAMPLE REVIEWS

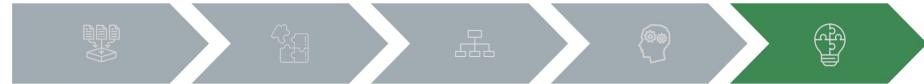
```
> sample(wpfpthin, 20)
[1] "t is now 5:00pm. the material is so thin it looks cheap. the knees bagg"
[2] "e. the material was strange- looked thin and cheap, and not at all what"
[3] " are fine, but the material is very thin compared to my older pants. th"
[4] "disappointing! thin material (way thinner than yea"
[5] " new ones is the material. it feels thin and clammy even after washing,"
[6] "ic could be thicker, it is a little thinner than i like so it does not hol"
[7] " the material on the stone color is thinner than the black pants. and ... "
[8] "ity of the fabric. it is noticeably thinner than it used to be. i own 6 pa"
[9] "at i previously ordered. it is much thinner and cheaper appeaering. i cons"
[10] "ore for an apple shaped person with thin legs. whatever happened to a f"
[11] "er fabric and sizing. the fabric is thinner and does not wash well. the pe"
[12] "r winter are not the same material. thinner and cheaper looking, like some"
[13] "ers who presume all short women are thin (and of course we all sew -- o"
[14] "to cold weather. the fabric is much thinner that was previously used. i do"
[15] "her than all that, the fabric seems thinner than original fit but the pant"
[16] " them all the time. these were of a thinner fabric, too high in the waist."
[17] "t fit leggings\"! also the fabric is thinner and flimsier than it used to b"
[18] "or years. each time the fabric gets thinner and they just don't hang write"
[19] "ime were substandard. they are very thin and did not hold their form af"
[20] "perfect fit pants, bootcut thin fabric i just tried on my new "
```

CONCLUSION



Photo from: flickr.com/photos/missnatalie/

- Demonstrated a classification algorithm to quantify customer sentiment based on review text
- Identified product improvement opportunities by text mining product reviews



NEXT STEPS

- Analyze larger dataset provided by L.L. Bean.
- Incorporate weighting factor for “Other users found helpful” field.
- Other product improvement opportunities.
- Increase granularity of review categories.



Photo: Instagram @llbean

THANK YOU!



Photo from: flickr.com/photos/ryan_roxx/