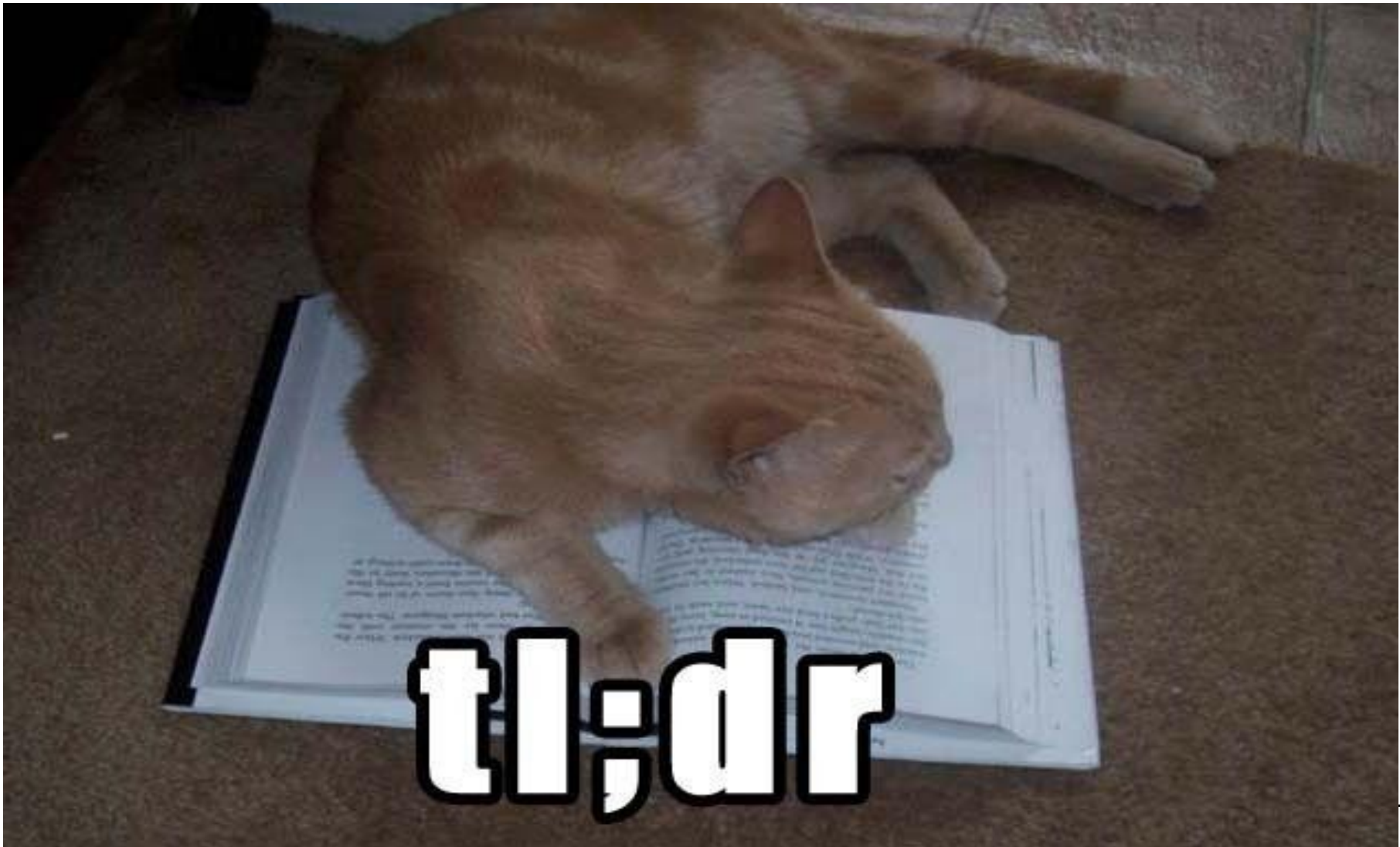# Review Filtering

Motivation

- Many reviews on a product/restaurant/Movie/book

- 100+ reviews are frequently seen

- But people stop after a few

- Find unique set of review

Data

- Amazon customer review on Major Applicance

- Year 2002-2015

- 96 k reviews,

- 90 k users, 11k products

- Too sparse for clustering

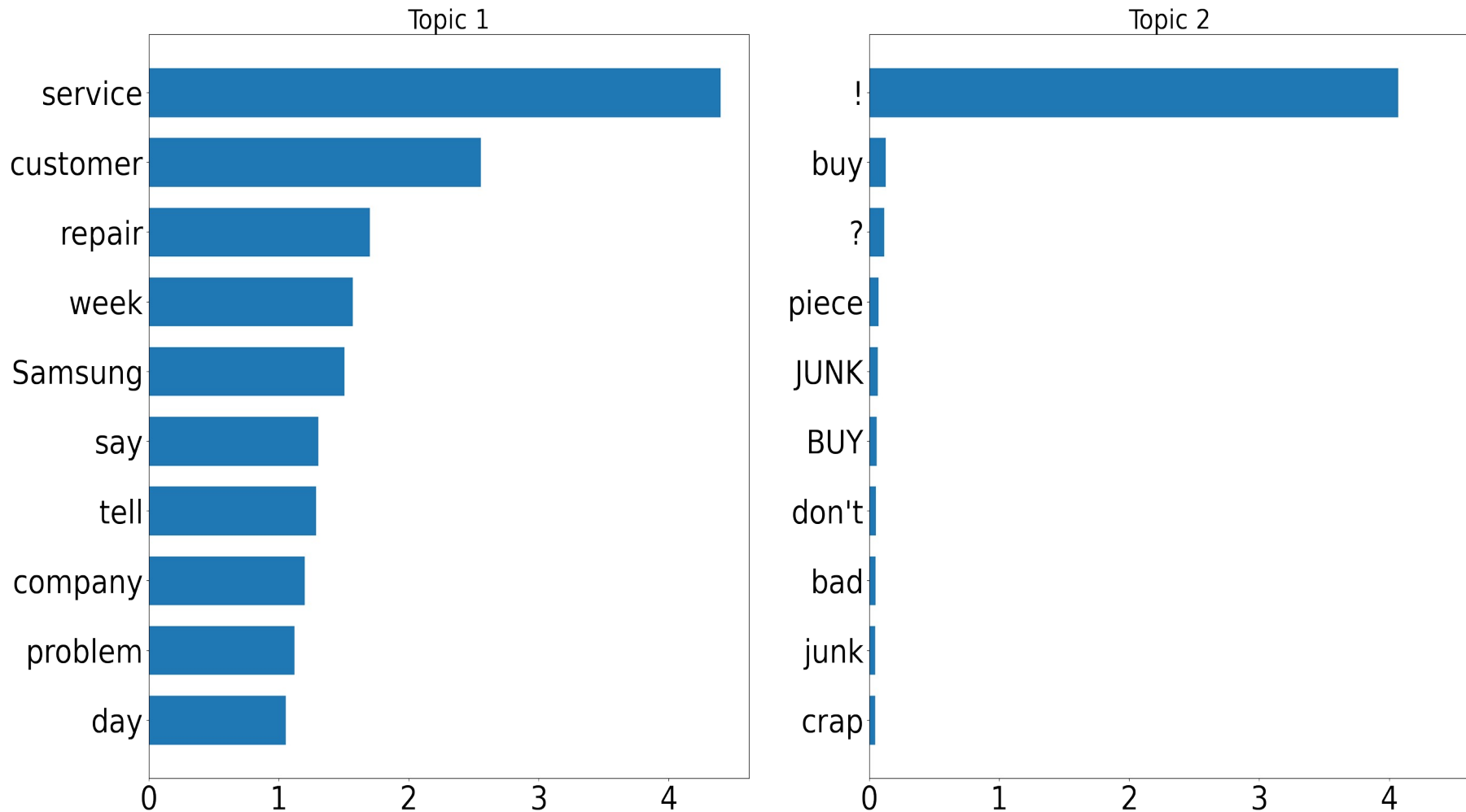# Pipeline

spaCy → **TF·IDF**

↓

COSINE SIMILARITY ← **NMF**

Summary

- Appliance Specific topic and Sentiment specific topic

- Some topics feel similar to humans, but treated as different topics by NLP algorithm, since words used are different .

- Some topics are dominated by "!" rather than words

Topics for Negative Review
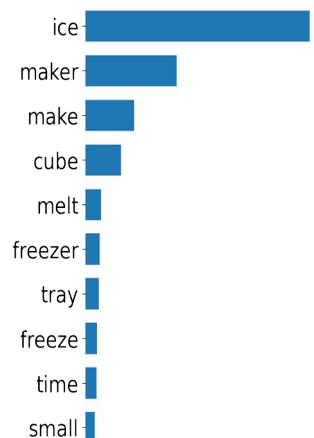
Top 2 topics for Negative Review

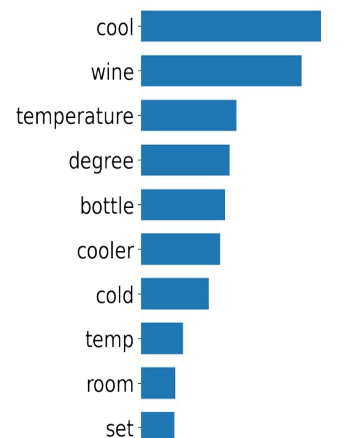# More Topics(Negative Review)

top topics for Negative Review
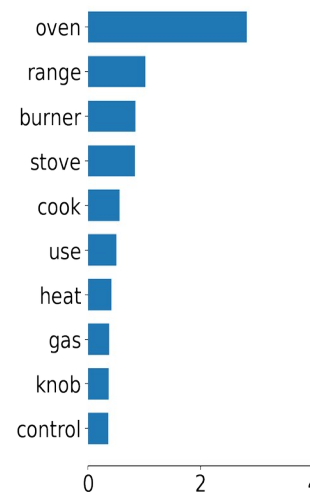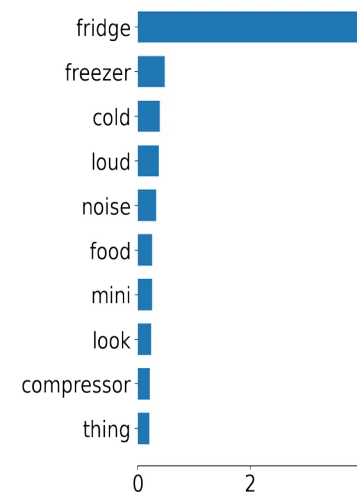
Topic for positive Review
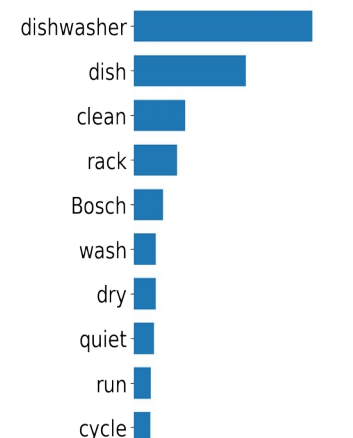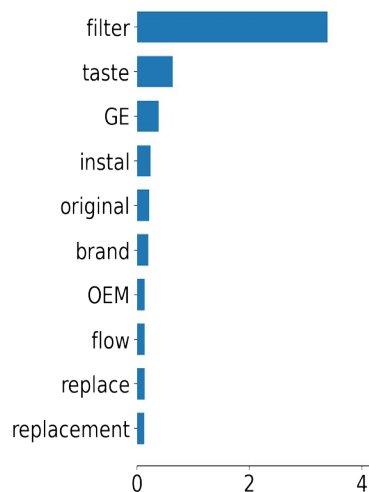
Top 2 topics for Positive Review
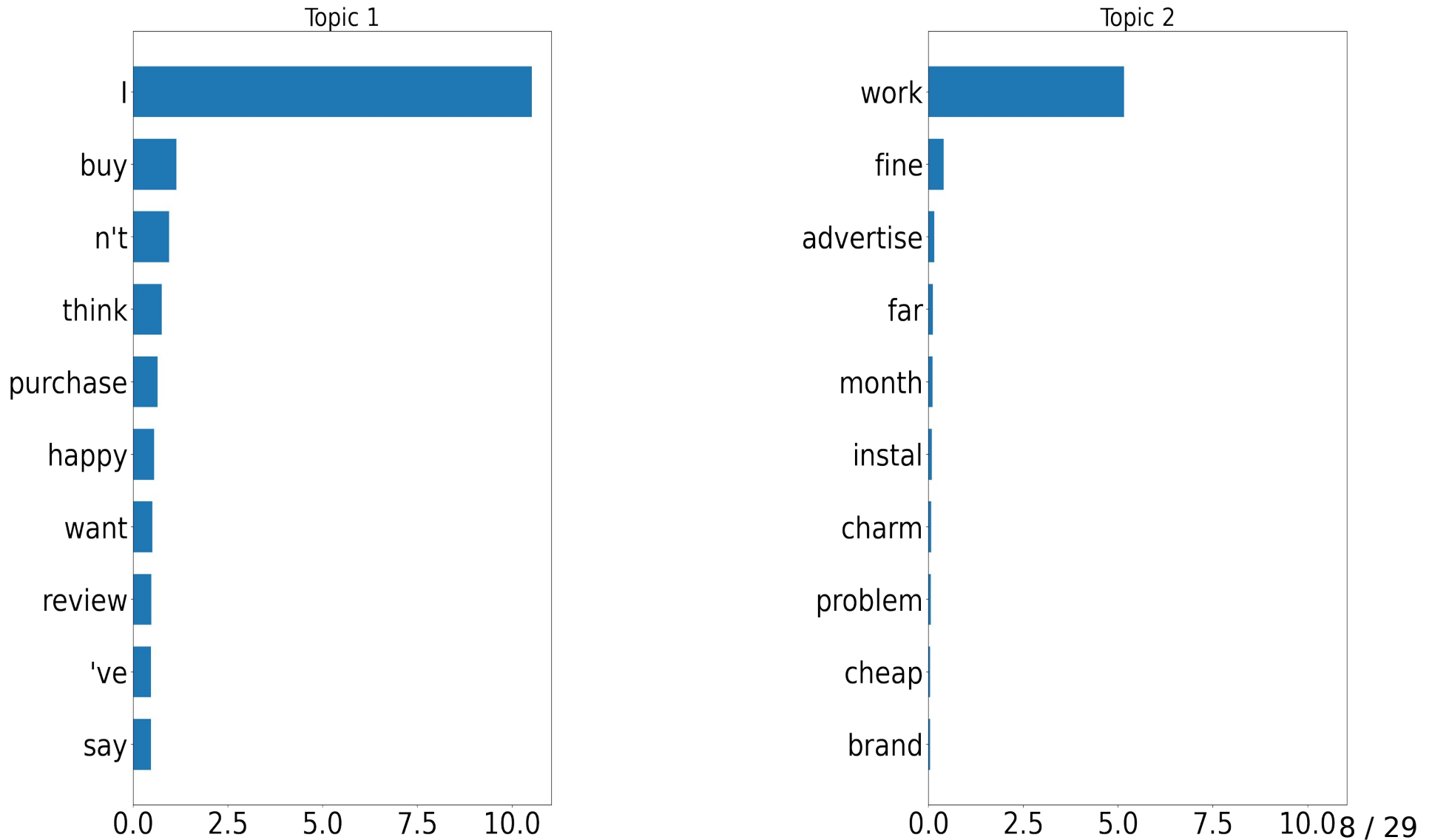
More Topics for Positive Review

top topics for Positive Review
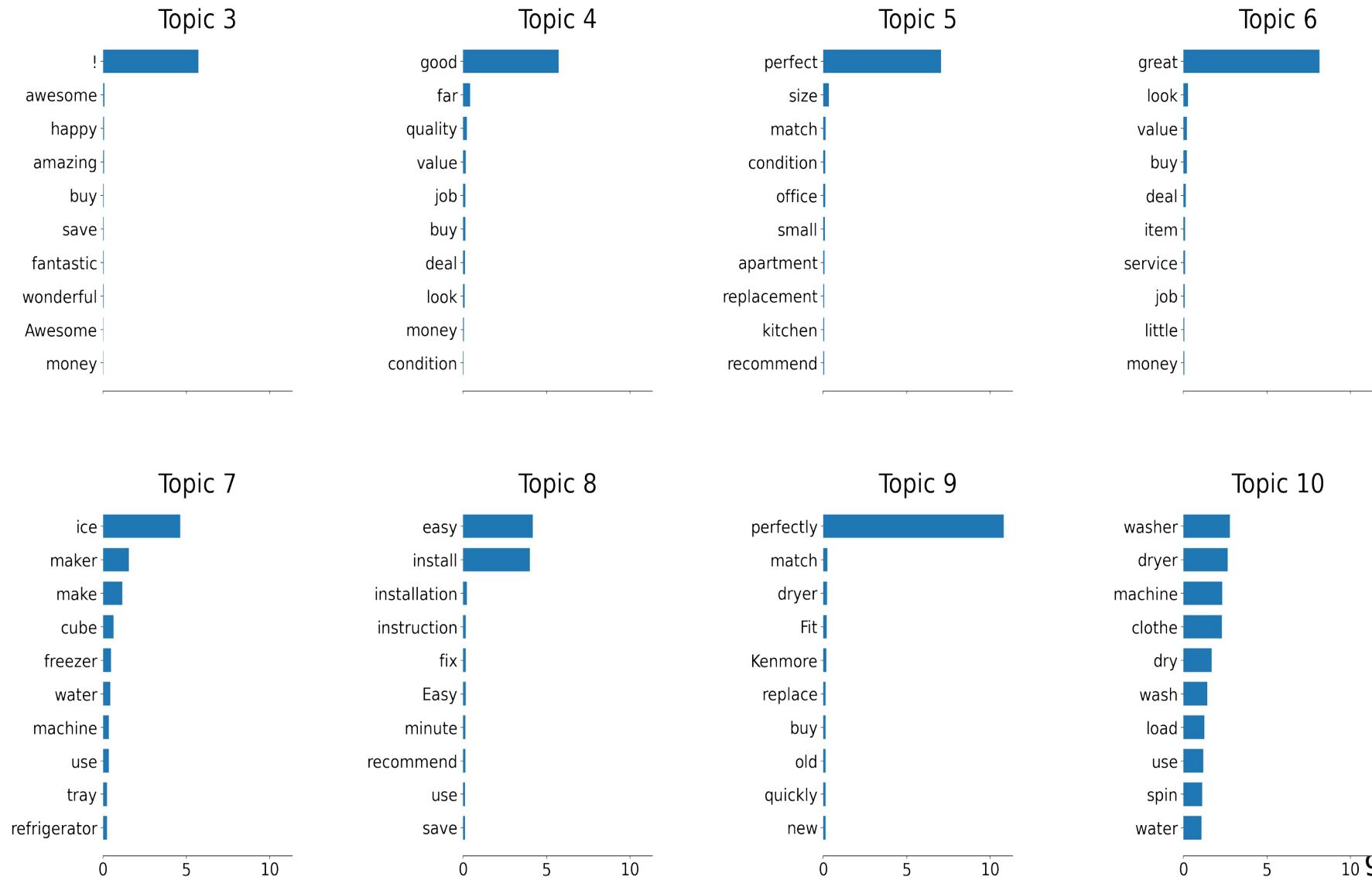
Filtering

- For each Product:

- Reviews with High pair wise cosine similarity (> 0.6) with any other unique reviews marked not-unique

Comments Reduction for Negative Review

- product   comments unique   diff   percent
- B000JLNBW4   190   30   160   0.842105
- B005KT4LK6   148   17   131   0.885135
- B00KJ07SEM   112   11   101   0.901786
- B002MPLYEW   121   29   92   0.760331
- B0058CLNBU   112   21   91   0.812500

Comments Reduction for Positive Review

- product_id   comments unique   diff percent
- B00KJ07SEM 1116   70   1046 0.937276
- B000JLNBW4 729   55   674 0.924554
- B004MX8XO6 601   52   549 0.913478
- B002C8HR9A 532   46   486 0.913534
- B002MPLYEW 540   55   485 0.89814

Future Work

- Smaller Corpus, from reviews on same type of appliance

- Add relationship between words of similar meaning

Conclusion

- Show unique reviews rather than all reviews

- Major reduction in clutter is possible

Appendix

Methodology

- Separate corpus for Positive / Negative Review

- Spacy for lemmentation

- Keep special punctuation, !, '

- Keep Upper Case letter

- TFIDF Vectorizer

- NMF for Topic modeling

- Cosine Similarity to find unique

Unique Versus Non-unique review

- Unique : I love this machine but it is so loud!! If your machine is in the garage or sits in the basement then it would be perfect. If your machine sits upstairs near bedrooms or close to your kitchen.... Forget it!

- Non-unique: I wish I could take this back. It doesn't clean. The clothes come out very wrinkled and full of lint in creases and folds. I used the detergent type suggested, and fabric softener, and even with all this, even T-shirts need to be ironed after the dryer. I can't wait until it breaks so I can get a different machine, and I've only had it 3 weeks.

Histogram of non-unique Negative reviews
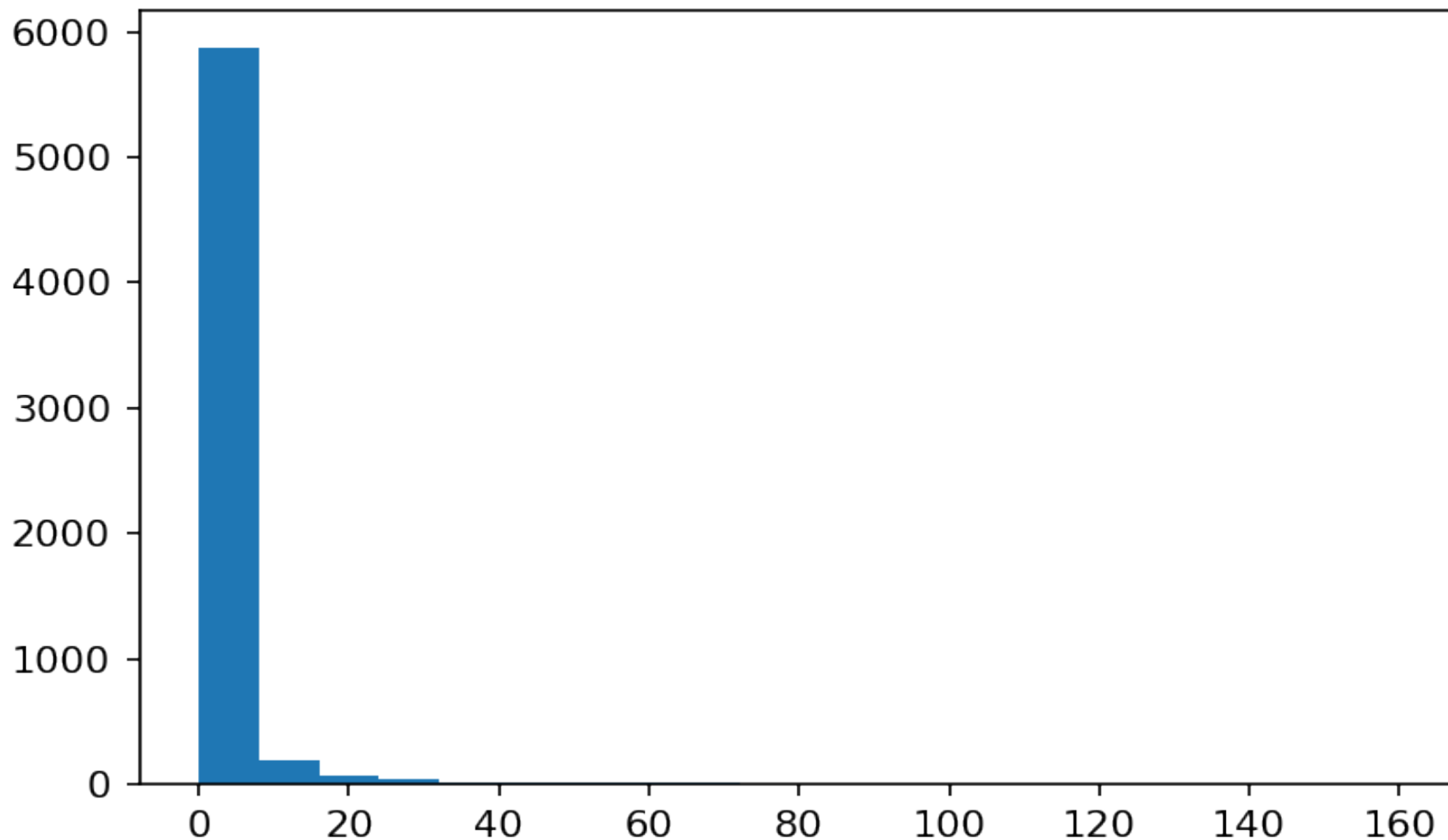
Non-unique Reviews

Histogram of non-unique Positive reviews
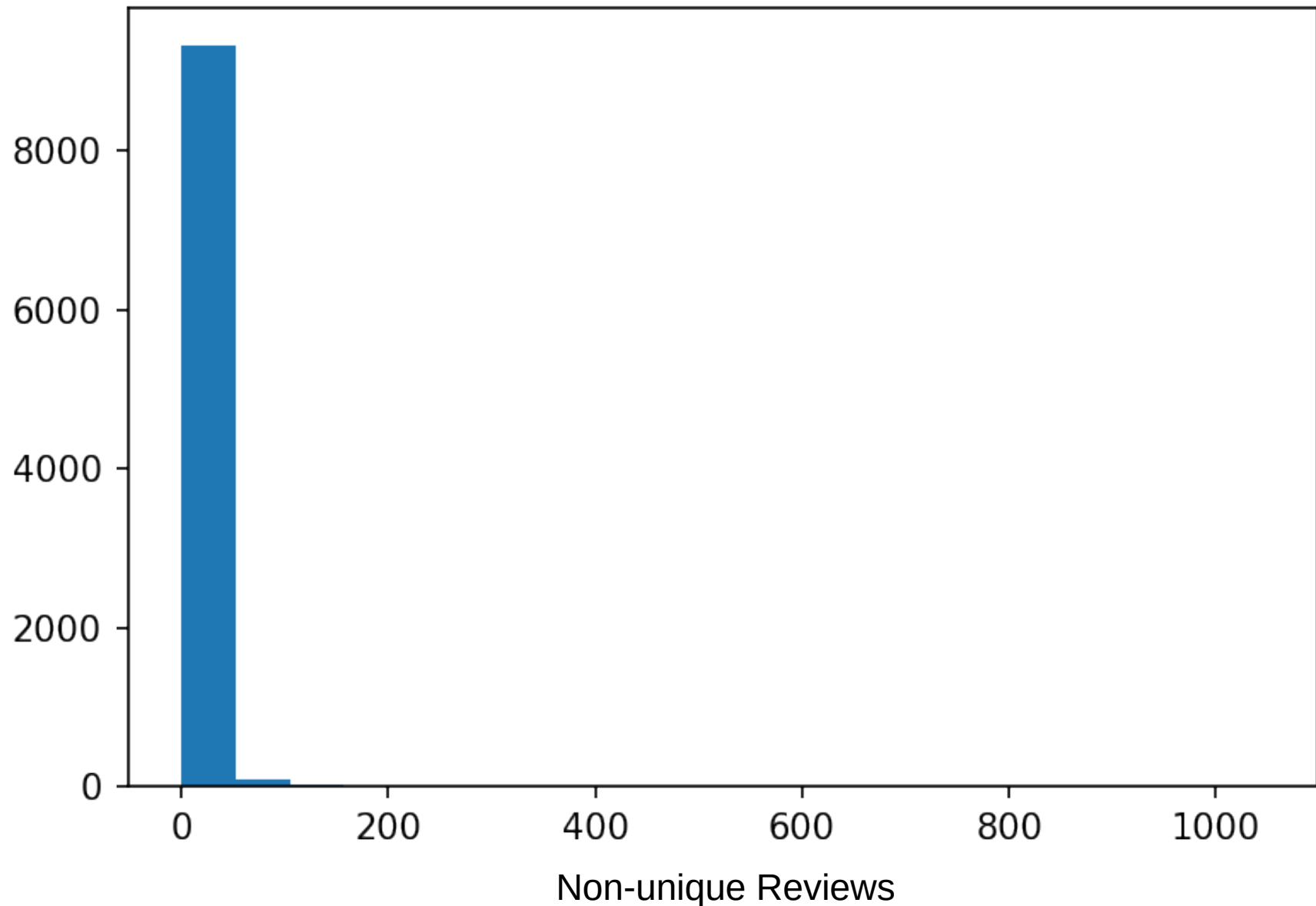
Remove ' in preprocessing

2 topics for Negative Review
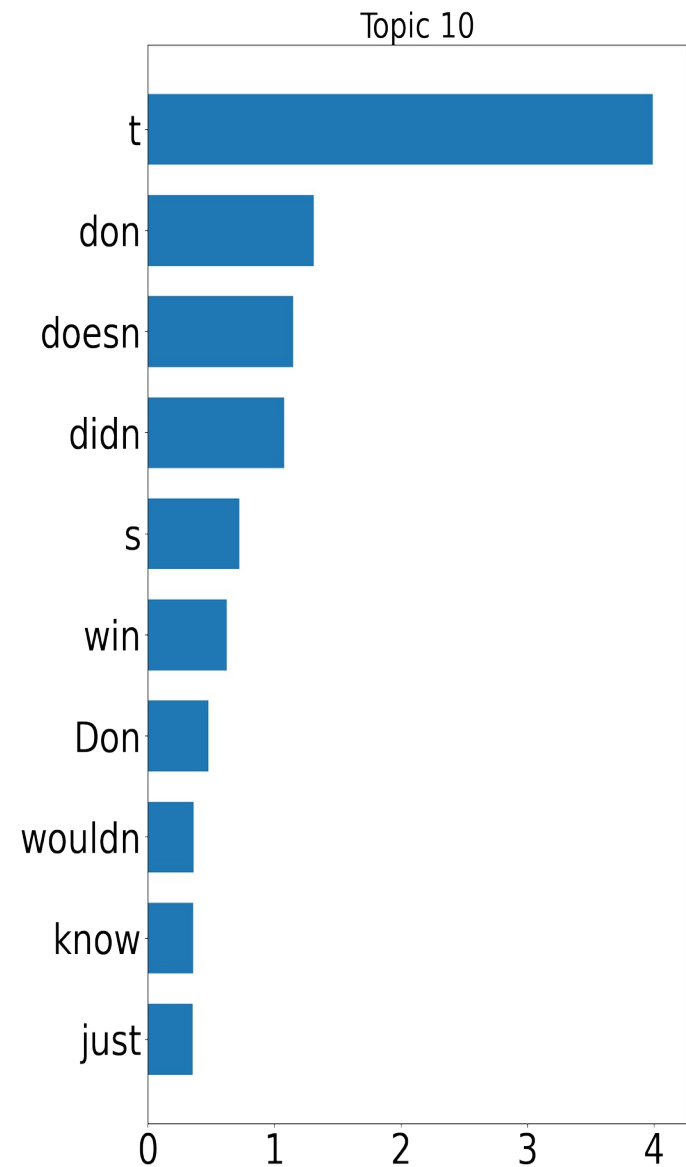
Positive Topics

- I buy n't think purchase happy want review 've say

- work fine advertise far month instal charm problem cheap brand

- ! awesome happy amazing buy save fantastic wonderful Awesome money

- good far quality value job buy deal look money condition

- perfect size match condition office small apartment replacement kitchen recommend

Positive Topics

- great look value buy deal item service job little money

- ice maker make cube freezer water machine use tray refrigerator

- easy install installation instruction fix Easy minute recommend use save

- perfectly match dryer Fit Kenmore replace buy old quickly new

- washer dryer machine clothe dry wash load use spin water

- product recommend happy highly satisfied

Positive Topics

- product recommend happy highly satisfied service quality awesome advertise pleased

- fit right space original nicely exact counter bottle plate tight

- love gift wife daughter absolutely husband kitchen size beautiful son

- filter water taste GE brand refrigerator flow expensive cheap OEM

- price right reasonable quality half beat store low delivery pay

Positive Topics

- just fine need right want size little original brand look

- fridge wine bottle cold cool temperature freezer door refrigerator shelf

- thank service fix new save order Amazon quick promise receive

- excellent service quality value condition delivery recommend job customer highly

- nice look really unit size stove quiet compact feature money

- [5.87719729 0.54548356 0.26870979

Positive Topics

- dishwasher dish clean n't use rack wash run come basket

- exactly need want I fix dryer replace advertise order repair

- fast shipping delivery ship quick super service item quality shipment

- expect exactly small early soon ok perform quality deliver little

- like charm new really original look champ brand suppose glove

- replacement original exact replace old break

Negative Topics

- n't ca wo buy know just 've thing cold think

-    ?   say cost * tell pay model know

- door open close handle design fall seal shelf freezer hinge

- year buy warranty repair replace cost ago old just new

- microwave GE heat food appliance new magnetron model turn keypad

Negative Topics

- machine washing Maytag drain pump purchase problem hose load new

- product buy quality recommend purchase bad stand poor company review

- Samsung drum issue problem appliance repair crack pump buy fail

- refrigerator freezer food GE problem freeze drawer purchase compressor temperature

- use time clean day cooktop burner small week light try

Negative Topics

- month warranty die stop start replace later purchase buy fail

- return item damage box Amazon receive dent arrive order shipping

- like look make noise loud good fan just really ?

- it bottle small stove shelf size wine space model knob

- break plastic handle cheap piece junk week rack quality replacement

Negative Topics

- LG repair lg appliance say problem come refuse technician buy

- dryer dry drum lint element heat crack belt set heating

- money waste time worth save junk spend piece nt total

- washer Whirlpool spin bad load Maytag new pump purchase problem

- water leak taste dispenser floor hot cold drain pump gallon