



Help Consumers Find Wines They'll Love

Using Natural Language Processing



Intro

Customer gives description of what kind of wine they love

set of 150K reviews from Wine Enthusiast

Find wines from the set of reviews that match with customer preferences



Method

Vectorize text of reviews into term matrix

Reduce from 4,000 words to 300 topics using
Non-Negative Matrix Factorization

Convert user description into 300-topic matrix

Find reviews that match user description best
by cosine similarity

Suggest best 2 matches

Topics

Dimensionality reduction produced meaningful topics

Some are general to all wines, and not helpful in selecting a wine for a customer

Topic 0

tannins, soft tannins, dry tannins, astringent, smooth tannins

Topic 1

acidity, crisp acidity, fresh acidity, bright acidity, brisk

Topic 2

fruit, dark fruit, passion fruit, passion, blackberry fruit

Topic 3

finish, finish drink, short, long finish, turns

Topic 4

red, red cherry, red fruits, red plum, red currant

Topic 5

cherry, blackberry cherry, red cherry, cherry blackberry, cherry berry

Results

Test efficacy of the model by
passing reviews of wines not in
the corpus

If the model suggests wines
similar to the test wine, the model
is working



Test 1:

Hyde de Villaine Chardonnay 2018:

"Opulent aromas of cooked apple lemon rind light praline and stone. Full-bodied with linear and lively acidity that runs through the wine. Energetic and bright with tightness and structure. Steely. Drinkable now but better in 2023."

['Louis Latour Pinot Noir Les Chaillots Premier Cru',

'Pascal Doquet Chardonnay Le Mesnil sur Oger Grand Cru Blanc de Blancs Brut']

Test 2:

Altesino Brunello di Montalcino 2017

"Inviting aromas of ripe plum violet baking spice and leather waft out of the glass. Full-bodied and dense the approachable palate features fleshy black cherry star anise and mocha while velvety tannins provide gentle support. It's already ready so drink soon."

['Masi Red Blend Costasera',

'Tenute Fiorebelli Red Blend Coste Bianche']

User input:

full-bodied with dark fruit, chocolate and coffee flavors”

Recommendations:

['Stark-Condé Cabernet Sauvignon Stark',

'Dr. Loosen Riesling Wehlener Sonnenuhr Auslese Gold Capsule']

Conclusion

Model makes a mix of relevant and non-relevant recommendations given text input

Draw on a larger corpus or reviews

Use more dimensions in the model

Refine stop words

Only focus on adjectives and nouns