

What are you saying?

**Introversion/Extroversion Algorithm: Mod 5 Project
Alaska Lam**

Problem Statement:

To better understand possible differences in online writing style between introverts and extroverts

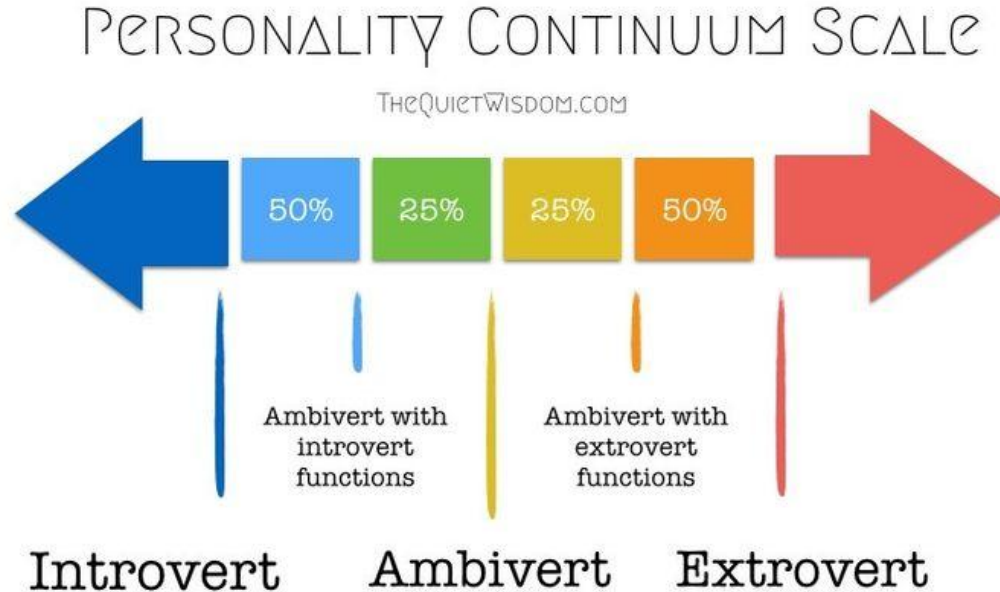
*writing conversationally about themselves.



Business Value:

Create foundation for more
comprehensive friendship-algorithm

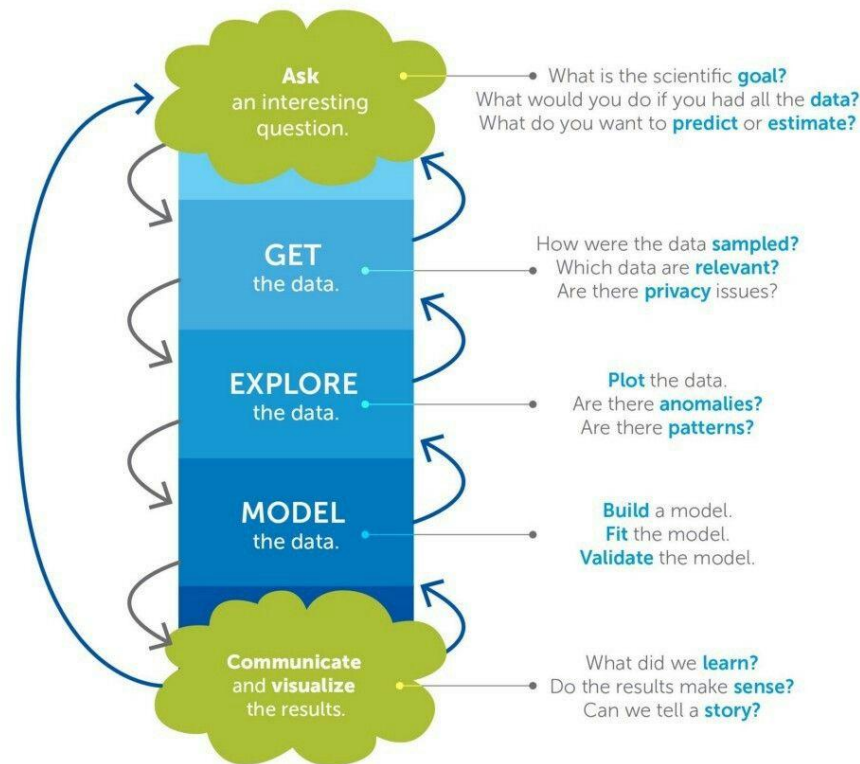
Capitalize on marketing demographics



Methodology:

Analyze past text-data written on Personality Cafe.net, a forum for talking about your life experiences

The Data Science Process



Derived from the work of Joe Blitzstein and Hanspeter Pfister, originally created for the Harvard data science course <http://cs109.org/>.

Model Results - Focus on Increasing F1 Score: Introversion vs Extroversion

F1 Score = Reduce
General
Mislabelling Errors

Category	F-1 Score [range from 0 to 1]
Introverts	.78
Extroverts	.51

Category	F-1 Score	Test accuracy
Everyone [introverts + extroverts]	.72 [weighted average]	.69

Recommendation 1, part 2:

Look at bigrams –

“!!” is double as common in extroverts

“me!” 1.5x for extroverts

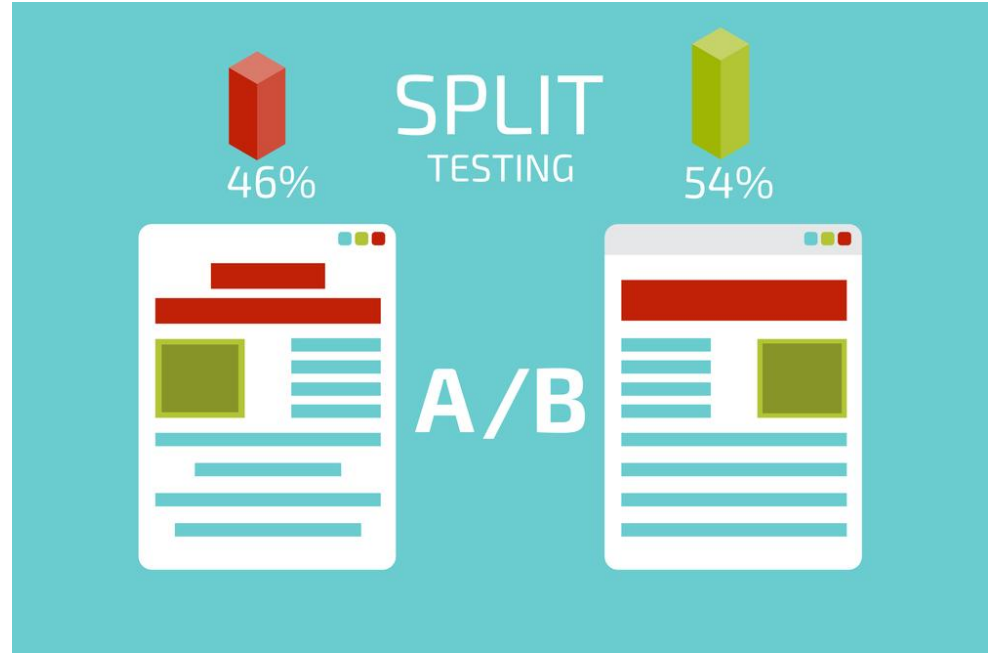
“too!” high on extroverts’ list



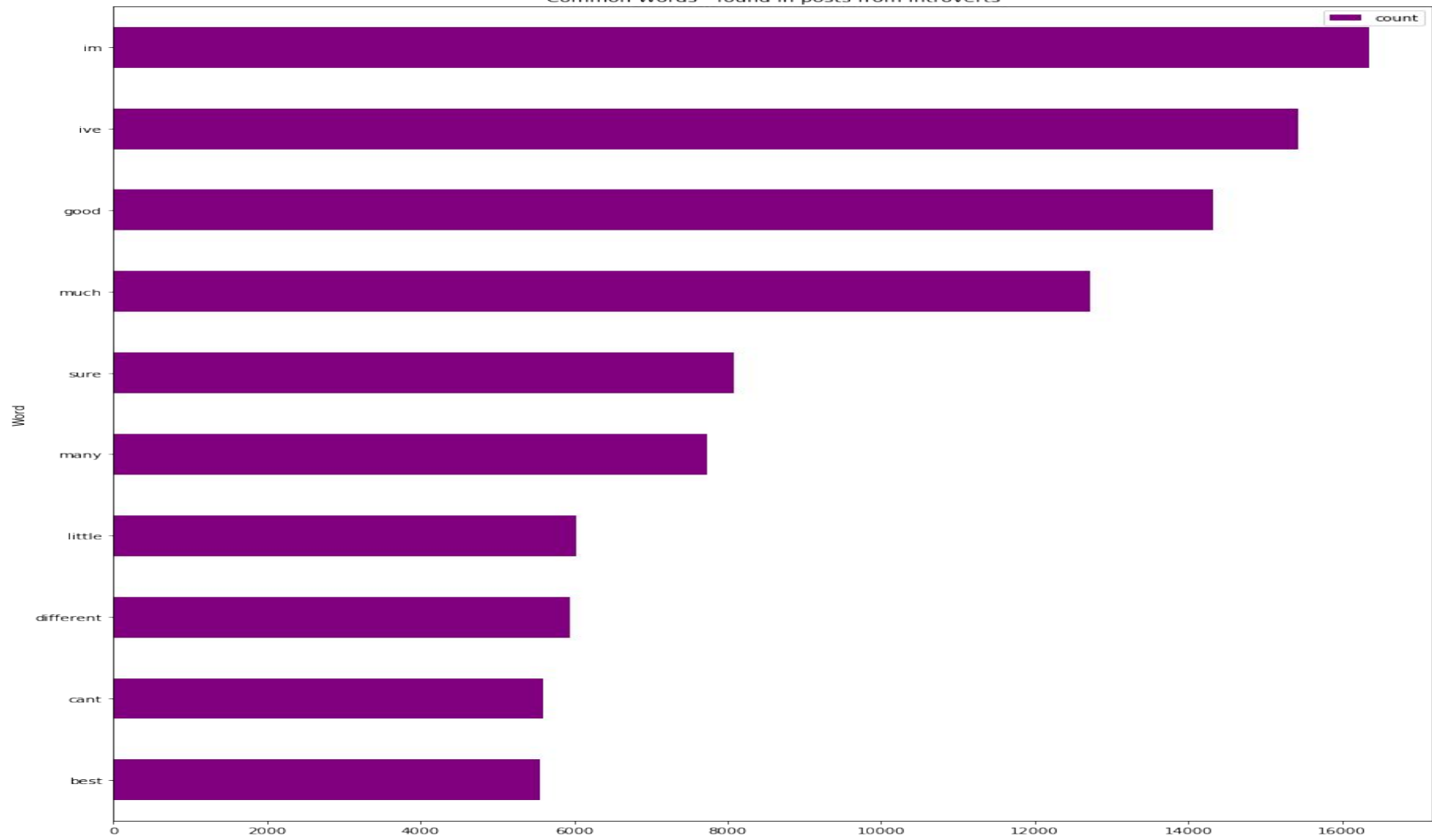
Recommendation 2:

Since all top words are similar when comparing introverts vs extroverts, do not attempt to create separate marketing campaigns at this point -

perhaps later for domain -specific algorithms



Common Words - found in posts from introverts



Recommendation 3:

Mutual information scores –
correlation with literary
references in introverts



Introverts - phrases that do not show up in extroverts' top pmi

giga blender - top one

whats subtype

monte cristo

jrr tolkien

forrest gump

los angeles

englishlanguage, arts

sodium,sodium

mockingbird,harper

bungee jumpingskydiving

dalai lama

winnie pooh

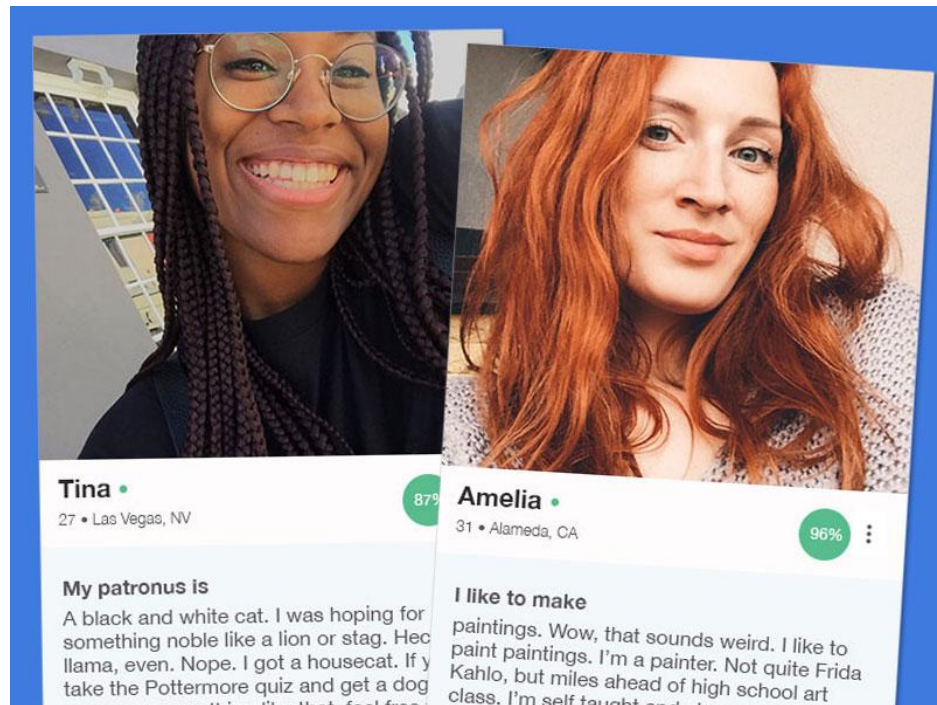
Future Considerations

Continue to fine-tune the current models, reduce overfitting

New model based on how often top bigrams or bigrams with top mutual info scores are used

Obtain more text-data from social events and websites such as OkCupid

Incorporate into friendship algorithm



Thank You!