Philippe Bernstoffe

Head of Marketing

Omega Group

Philippe is a newly hired head of marketing who has been tasked to **reclaim the #2 spot** in the Swiss Watch industry. His implementation strategy is to make more **data-driven marketing decisions**.

While most decisions can be evaluated with data, he has been **having trouble objectively assessing copy effectiveness** before launching campaigns. Different stakeholders coming from various backgrounds have their own agendas and subjective opinions.

He wants to be able to **confidently and objectively make decisions on advertisements before they are published**, and he believes that data-driven methods could pose an effective solution

For:

Ω OMEGA

Prepared by:

Clarence

Michael

Hongwei

Marketing with Data

Employing data science techniques to reduce subjectivity in ad creation

Agenda

01	Background
02	Problem Statement
03	Cleaning And EDA
04	Model
05	Stremlit Demo
06	Conclusion and Recommendation

Context:

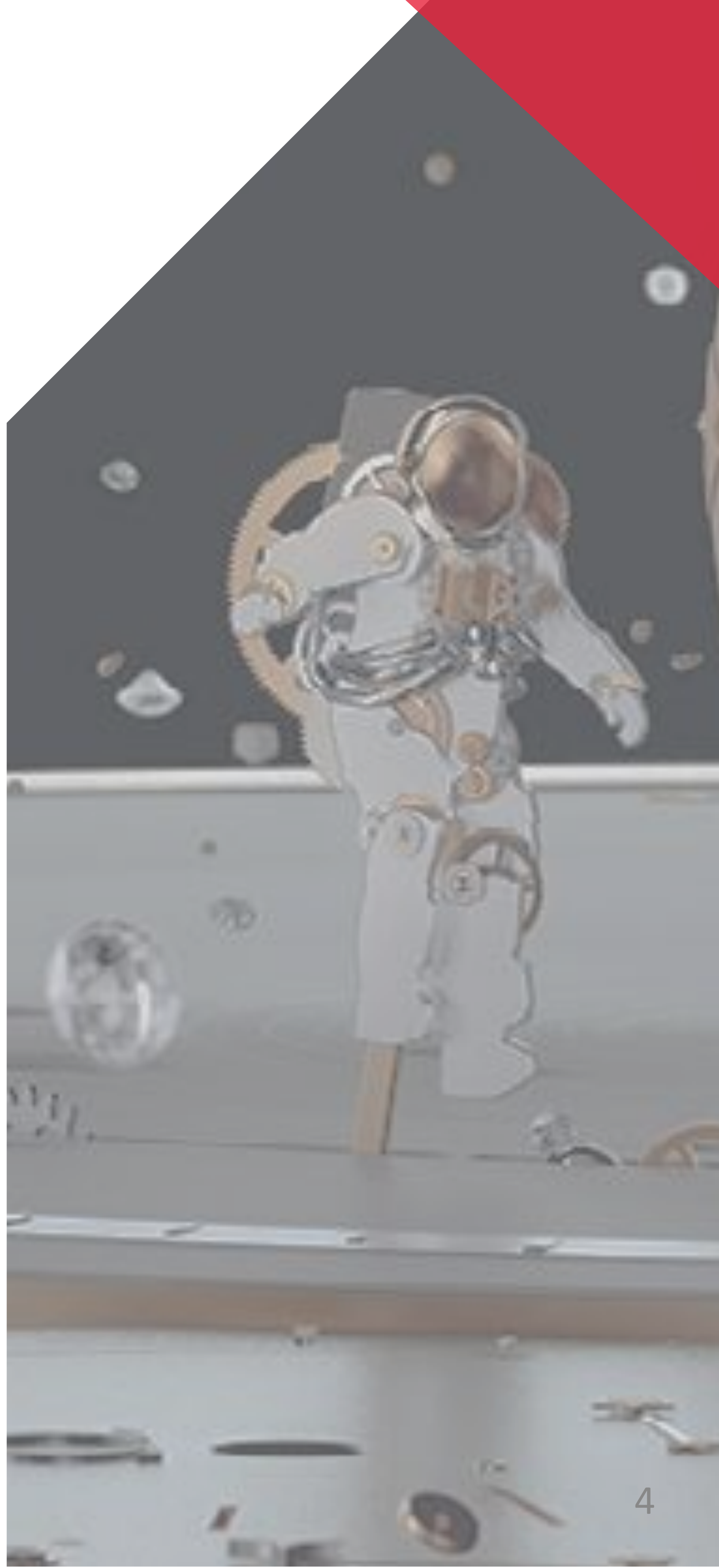
Data-Driven Marketing

Morgan Stanley's Top 20 Swiss Watch Company Ranking

- In 2020, Omega's longstanding #2 ranking was overtaken by Cartier
- Since then, Cartier's #2 position has been unchallenged
- Broad Marketing strategy for OMEGA is to make more data-driven decisions

Rank	2017	2018	2019	2020	2021	2022	2023
1	Rolex	Rolex	Rolex	Rolex	Rolex	Rolex	Rolex
2	Omega	Omega	Omega	Cartier Watches	Cartier Watches	Cartier Watches	Cartier Watches
3	Cartier Watches	Cartier Watches	Cartier Watches	Omega	Omega	Omega	Omega

source

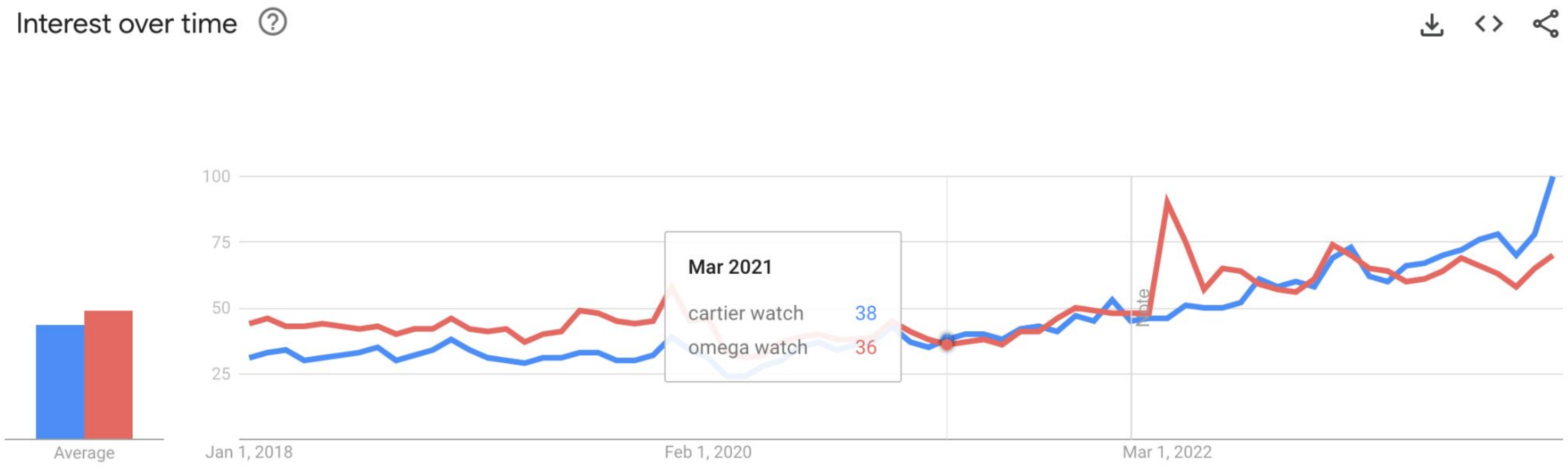


Context:

Cartier’s popularity rivaling OMEGA

Google search trends

- Momentary spikes for Omega, with searches for Cartier currently in the lead



source

Data guiding creativity

How data-driven can we be
when it comes to
evaluating language?

Guess the brand tagline

Live More, Bank Less

Just Do It.

I’m Lovin’ It.

It’s Finger Lickin’ Good

Gives you wings

Gotta Catch’em All

Guess the brand tagline

Live More, Bank Less

Just Do It.

I’m Lovin’ It.

It’s Finger Lickin’ Good

Gives you wings

Gotta Catch’em All

Guess the watch marketing slogan

Elegance in motion

Timeless sophistication

Precise in every detail

Timeless elegance,
defined by {product name}

Elevate your elegance with
{product name}

Love is All

Guess the watch ad campaign slogan

OMEGA

Elegance in motion - Constellation.

Timeless sophistication - De Ville

Precise in every detail - Aqua Terra

Cartier

Timeless elegance, defined by Tank

Elevate your elegance with Ballon Bleu

Cartier: Love is All (2022)

We get it

Precision?
Sophistication?
Elegance?
Exploration?

Which word sounds more OMEGA?

PRECISION!
Our customers care more about accuracy

Sales

EXPLORATION!
We have to be an aspirational brand

Branding

SOPHISTICATION!
We need people to know the craftsmanship behind the watches

Product

As is:

Subjective

Brand voice

Brands identify their voice and design ads through **non data-driven methods.**

This leads to subjectivity and **reduces accuracy in execution.**

Customer Research

Extracted by analyst

Keywords

Interpreted by
head copywriter

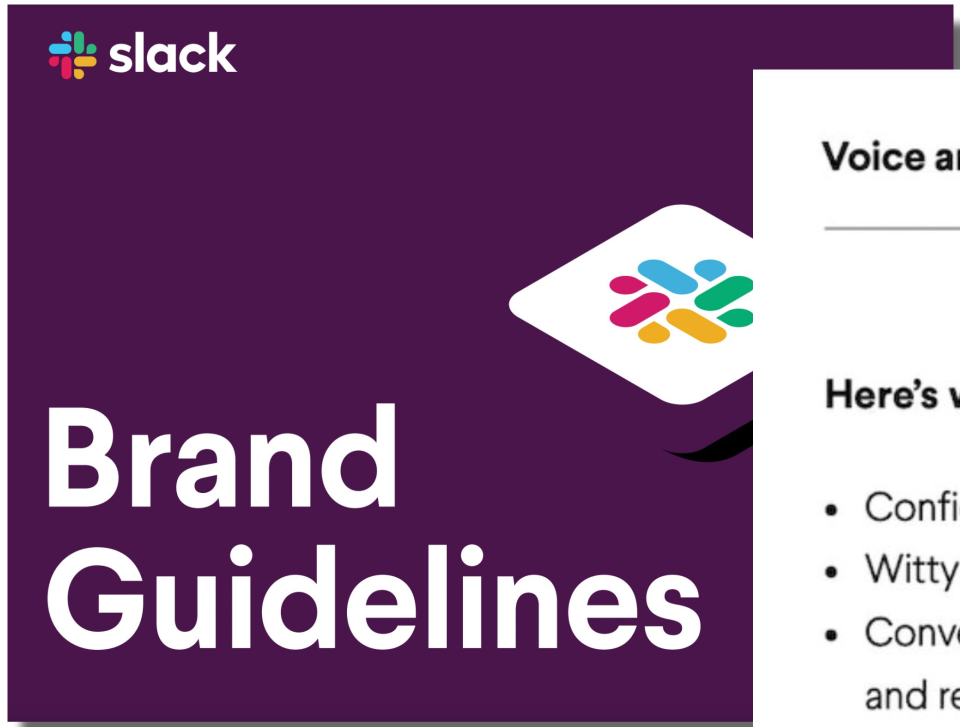
Voice Guide

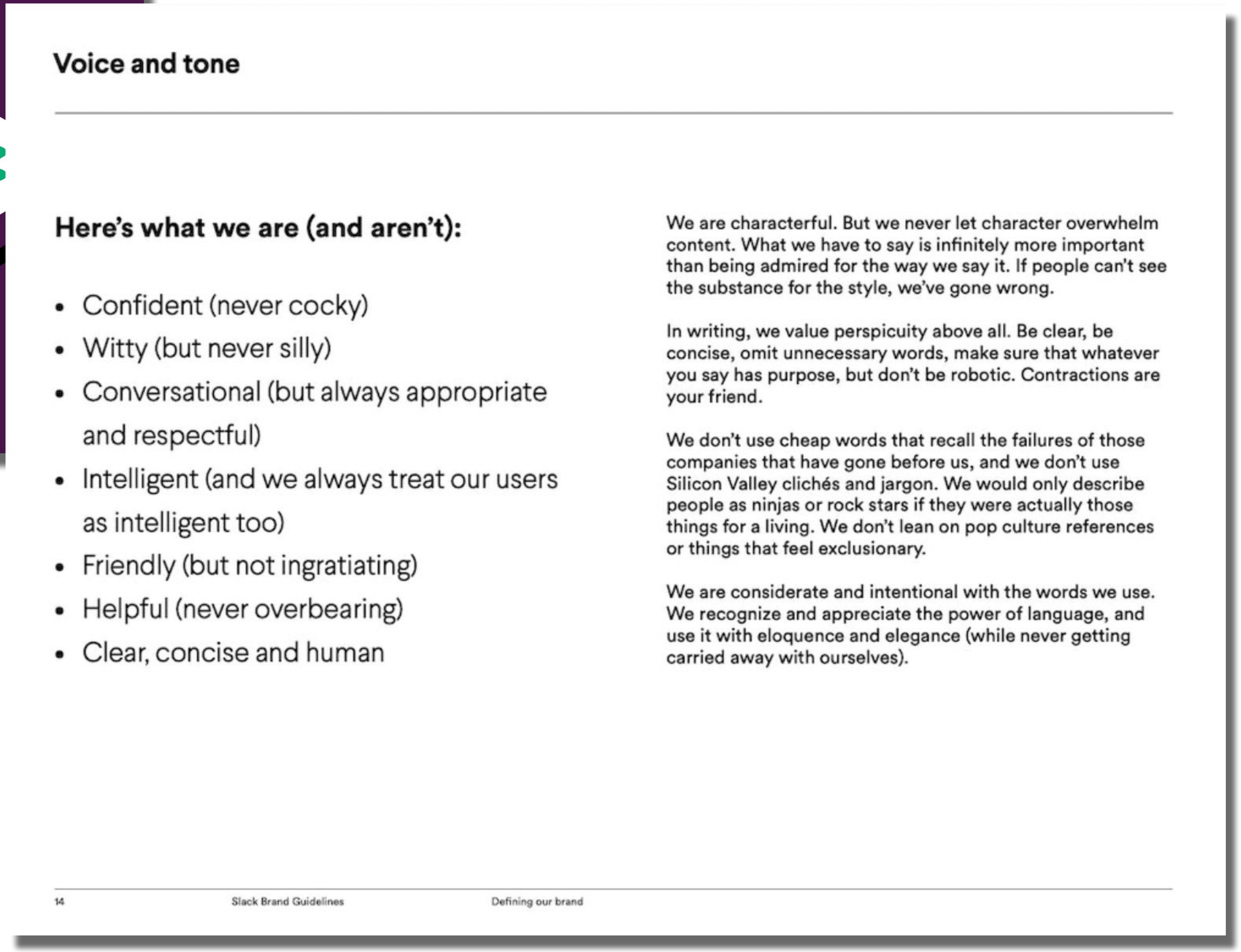
Interpreted by
junior copywriter

Advertisement

Example:

Brand voice guide





Voice and tone

Here's what we are (and aren't):

- Confident (never cocky)
- Witty (but never silly)
- Conversational (but always appropriate and respectful)
- Intelligent (and we always treat our users as intelligent too)
- Friendly (but not ingratiating)
- Helpful (never overbearing)
- Clear, concise and human

We are characterful. But we never let character overwhelm content. What we have to say is infinitely more important than being admired for the way we say it. If people can't see the substance for the style, we've gone wrong.

In writing, we value perspicuity above all. Be clear, be concise, omit unnecessary words, make sure that whatever you say has purpose, but don't be robotic. Contractions are your friend.

We don't use cheap words that recall the failures of those companies that have gone before us, and we don't use Silicon Valley clichés and jargon. We would only describe people as ninjas or rock stars if they were actually those things for a living. We don't lean on pop culture references or things that feel exclusionary.

We are considerate and intentional with the words we use. We recognize and appreciate the power of language, and use it with eloquence and elegance (while never getting carried away with ourselves).

14 Slack Brand Guidelines Defining our brand

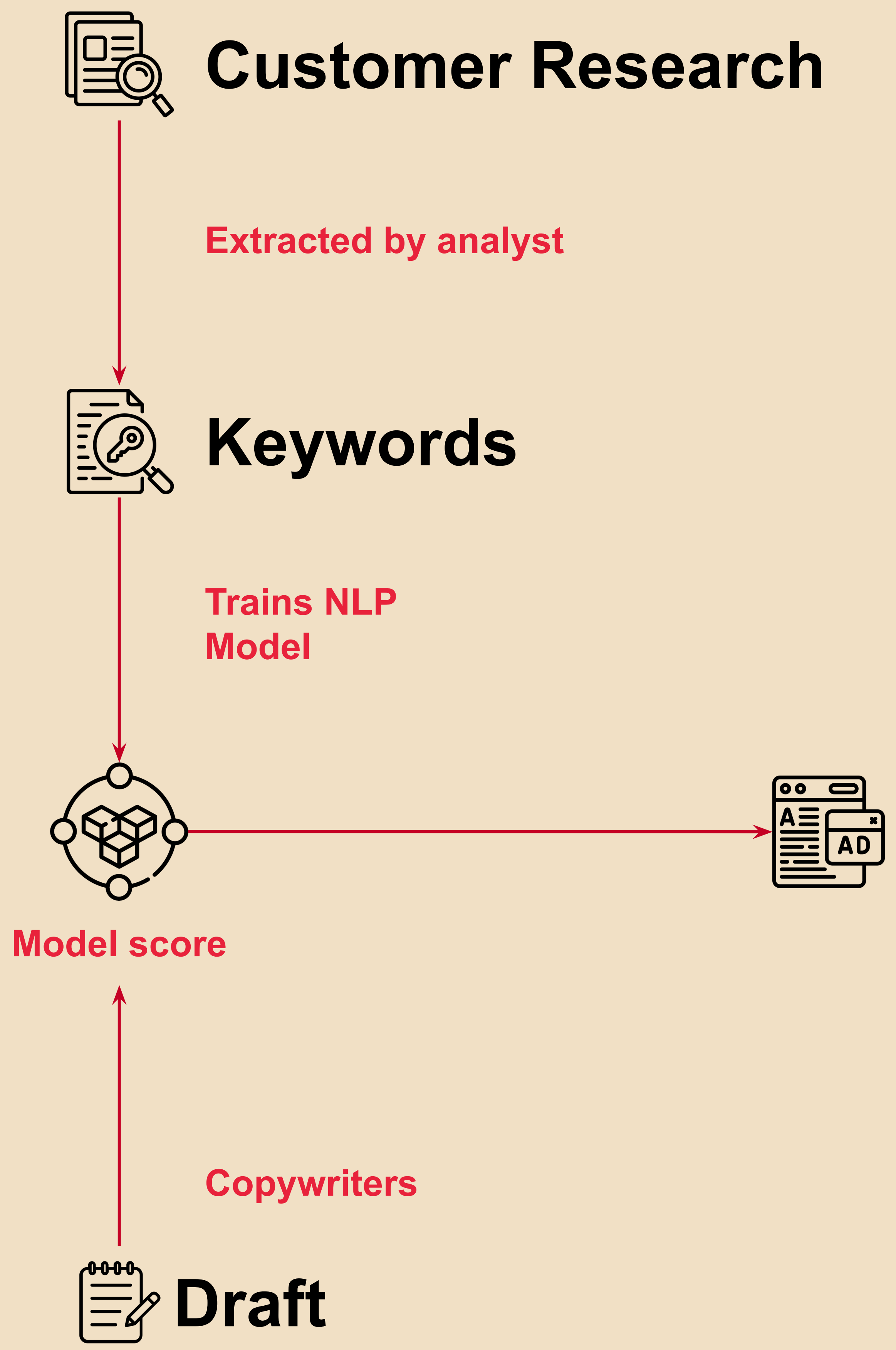
To be:

Data-driven

Brand voice ⇔ Customer voice

What if we could **skip all the subjective interpretation** of how good a word or phrase is?

How can we be **confident in proposing** a new **marketing line that resonates** with our customers?



The North Star & Barrier

Data-driven marketing decisions will help us reclaim #2 in industry ranking, it guides organisational objectivity and standard procedures.

Objectivity remains the plague of branding and advertising worldwide.

The Problem Statement

- Data-driven marketing decisions will help us reclaim #2 in industry ranking, it guides organisational objectivity and standard procedures.
- Objectivity remains the plague of branding and advertising worldwide.
- How might we leverage natural language processing to inform our marketing decisions with confidence and objectivity, especially on copy that goes to brand taglines and marketing slogans?

Methodology

Web Scraping → Cleaning & Preprocessing → Analysis → Modelling

Web Scraping

```
forum-page-thread-items">

        js-inlineModContainer
        js-threadListItem-190533" data-author="john wilson" qid="thread-item">
    ▶<div class="structItem-cell avatar-cell">…</div>
    ▼<div class="structItem-cell structItem-cell--main" data-xf-init="touch-proxy">
      ▶<h3 class="structItem-title">…</h3> == $0
      ▶<div class="structItem-minor">…</div>
     </div>
    ▶<div class="structItem-cell forum-view-stats-listing-price regular-post">…</div>
    ▶<div class="structItem-cell last-post-cell" qid="thread-item-last-post">…</div>
   </div>
  </div>
 ▶<div class="california-thread-item  " qid="thread-item-parent">…</div>
 ▶<div class="california-thread-item  " qid="thread-item-parent">…</div>
 ▶<div class="california-thread-item  " qid="thread-item-parent">…</div>
 ▶<div class="california-thread-item  " qid="thread-item-parent">…</div>
  <hr class="sticky-divider" qid="forum-page-sticky-divider">
 ▶<div class="california-thread-item  " qid="thread-item-parent">…</div>
 ▶<div class="california-banner-ad-container california-ad hide-mw-responsiveWide california-ad-in-
 ▶<div class="california-banner-ad-container california-ad california-ad-mobile california-ad-in
 ▶<div class="california-thread-item  " qid="thread-item-parent">…</div>
```

Data Collection

- **Where?** Watchuseek watch forum
- **Why?** Active community of watch enthusiasts

  Segmentation of different watch brands in sub-forums which facilitate web scraping process
- **How?** In-house webscraping
- **When?** 2011-2024

Back to Agenda

Data Collection

Data Standardisation

- Select posts that have at least 2 words to make meaningful insights
- 14,000 posts from each brand's sub-forum

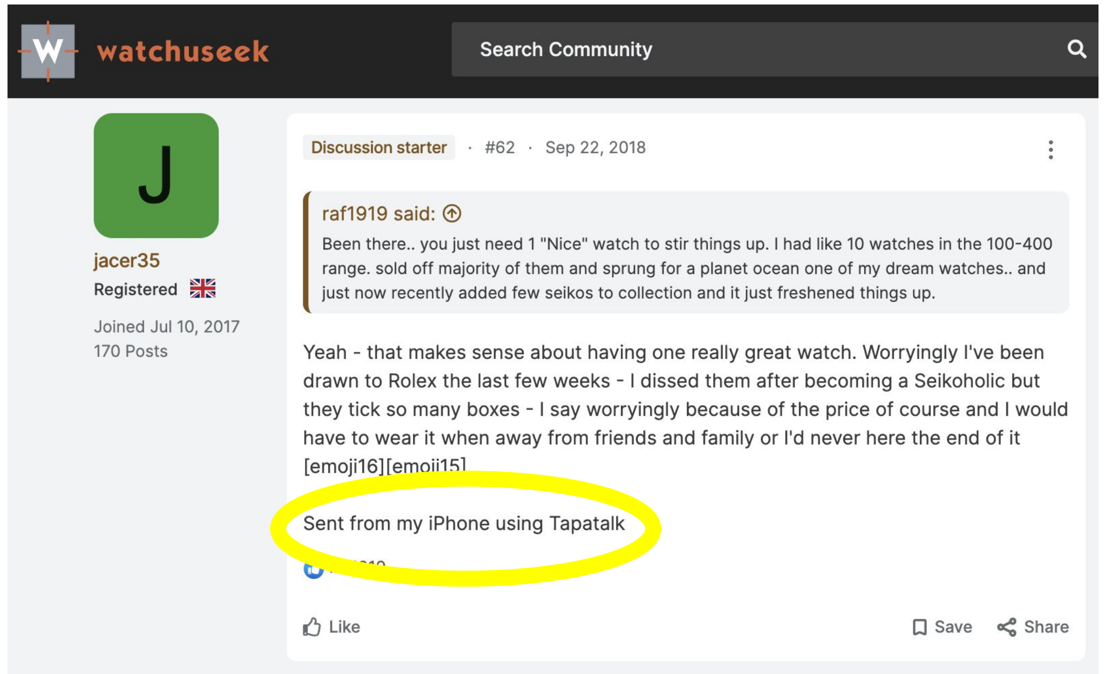
Features Collected

- 'author': The name of the author
- 'body': The content of the post
- 'date': The date the post was created
- 'like': 'True' if another user liked the post

Back to Agenda

Cleaning and Preprocessing

Data Cleaning



Removed:
- ‘http’, ‘www’, ‘com’, ‘watchuseek’
- Newline: '\n'
- punctuations
- ‘tapatalk’, ‘iphone’, ‘samsung’
- stopwords

```
> stopwords("english")
  [1] "i"          "me"         "my"         "myself"     "we"
  [6] "our"        "ours"       "ourselves"  "you"        "your"
 [11] "yours"      "yourself"   "yourselves" "he"         "him"
 [16] "his"        "himself"    "she"        "her"        "hers"
 [21] "herself"    "it"         "its"        "itself"     "they"
 [26] "them"       "their"      "theirs"     "themselves" "what"
 [31] "which"      "who"        "whom"       "this"       "that"
 [36] "these"      "those"      "am"         "is"         "are"
 [41] "was"        "were"       "be"         "been"       "being"
 [46] "have"       "has"        "had"        "having"     "do"
```

Data Preprocessing

Building → Build
Tester → Test
Engineered → Engineer

$\xrightarrow{\textit{Lemmatization}}$

Analysis

Initial Findings

Back to Agenda

Initial Findings

Omega Top 10 words

1. ~~watch~~
2. omega
3. ~~one~~
4. ~~look~~
5. bracelet
6. ~~strap~~
7. ~~get~~
8. dial
9. time
10. new

Cartier Top 10 words

1. ~~watch~~
2. cartier
3. ~~one~~
4. tank
5. ~~look~~
6. santos
7. ~~strap~~
8. think
9. movement
10. ~~get~~

Back to Agenda

Initial Findings

Omega Top 10 word pairs

1. rubber strap
2. planet ocean
3. aqua terra
4. look great
5. year ago
6. good luck
7. seamaster 300
8. omega seamaster
9. end link
10. great watch

Cartier Top 10 word pairs

1. cartier tank
2. de cartier
3. tank solo
4. cartier santos
5. leather strap
6. tank must
7. two tone
8. must de
9. dress watch
10. stainless steel

Back to Agenda

Initial Findings

Omega Top 5 trigrams

1. seamaster diver 300m
2. oem rubber strap
3. omega service center
4. second per day
5. flat link bracelet?

Cartier Top 5 trigrams

1. tank solo xl
2. tank solo large
3. cartier tank solo
4. tank louis cartier
5. cartier stepped tank

Initial Findings

Omega Top 5 trigrams

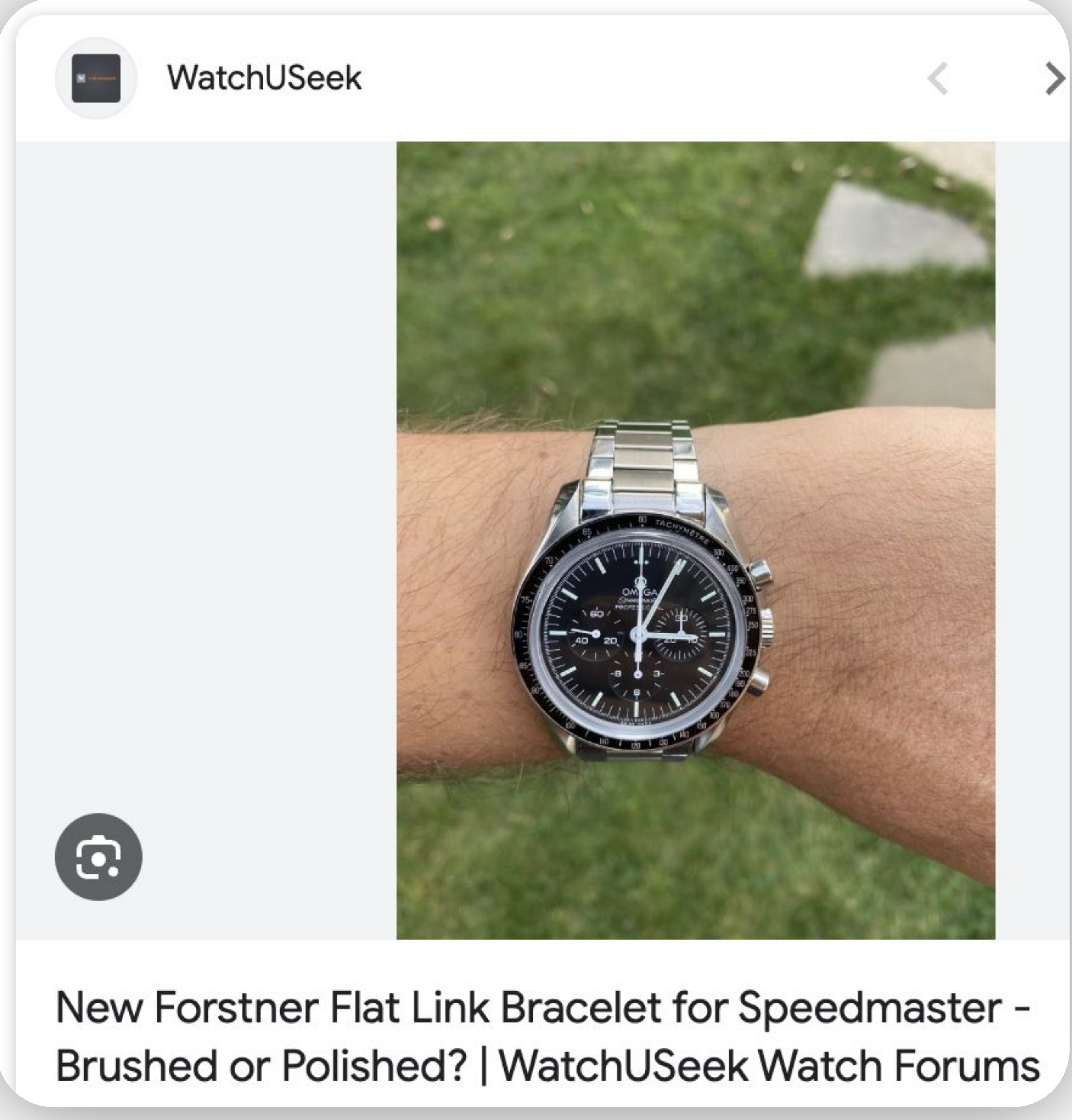
1. seamaster diver 300m
2. oem rubber strap
3. omega service center
4. second per day
5. flat link bracelet

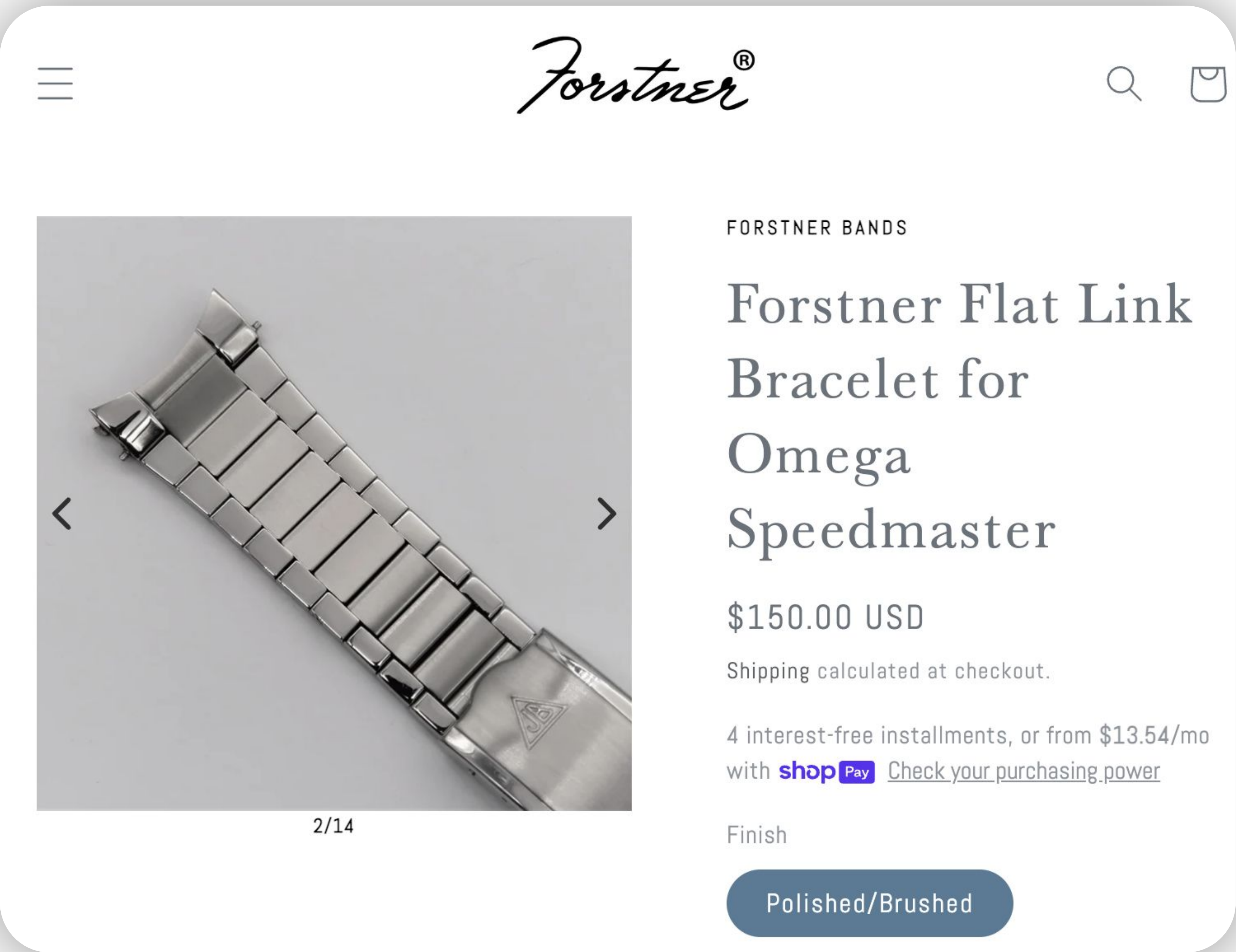
Cartier Top 5 trigrams

1. tank solo xl
2. tank solo large
3. cartier tank solo
4. tank louis cartier
5. cartier stepped tank

Back to Agenda

Initial Findings





Back to Agenda

Initial Findings

Omega Top 10 trigrams

1. seamaster diver 300m
2. oem rubber strap
3. omega service center
4. second per day
5. flat link bracelet
6. 20 01 001
7. omega aqua terra
8. master control date
9. official omega website
10. online official omega

Cartier Top 10 trigrams

1. tank solo xl
2. tank solo large
3. cartier tank solo
4. tank louis cartier
5. cartier stepped tank
6. cartier tank watch
7. cartier santos 100
8. cartier santos galbee
9. cartier tank louis
10. vintage cartier tank

Back to Agenda

Modelling

Transformer

Count Vectorizer vs Term Frequency Inverse Document Frequency Vectorizer

Omega is very very good

$$TF \times IDF = \frac{\text{number of the word appears}}{\text{total number of words}} \times \left(\log\left(\frac{\text{number of sentences}}{\text{number sentences that contain the word}}\right) + 1\right)$$

Omega	is	very	good
1	1	2	1

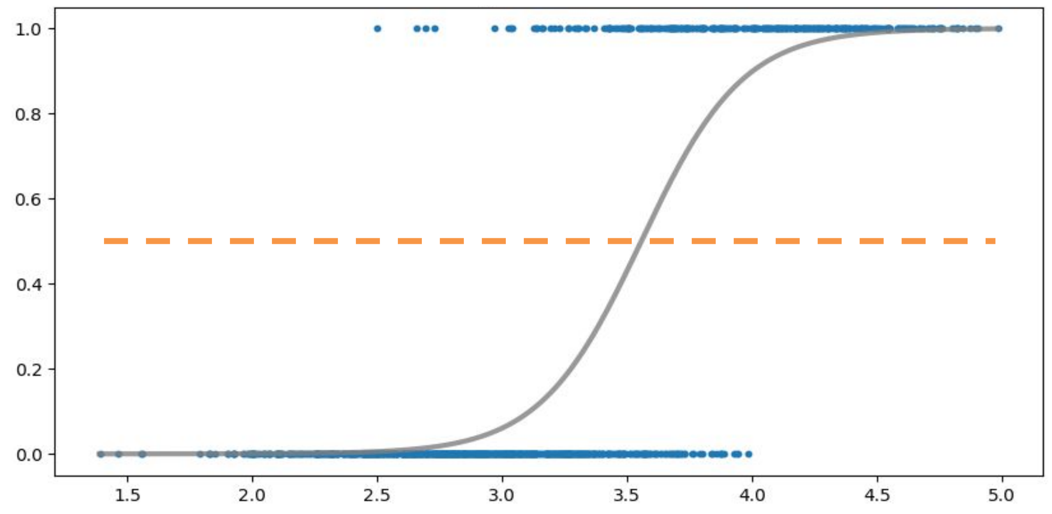
Count Vectorizer

Omega	is	very	good
0.2	0.2	0.4	0.2

TF-IDF Vectorizer

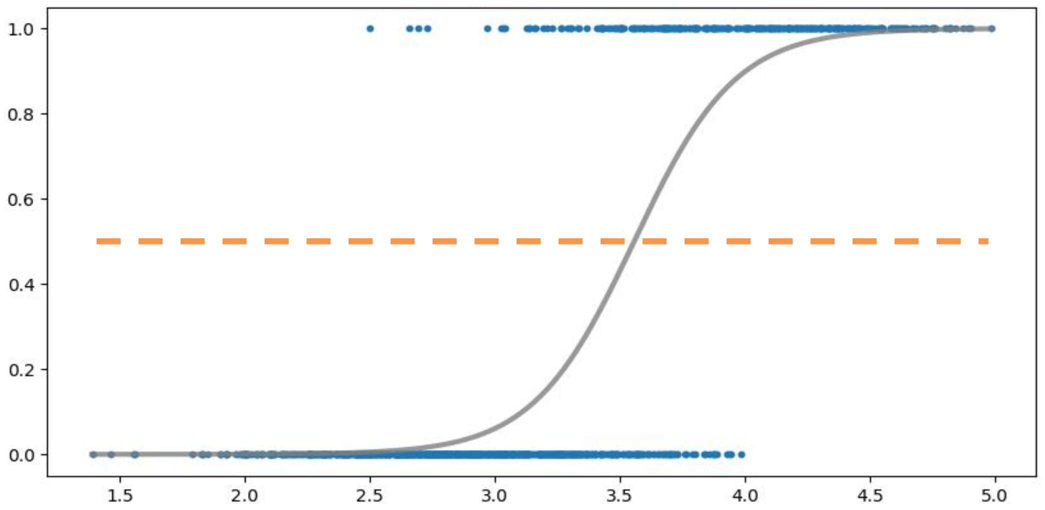
Estimator

Logistic Regression



Estimator

Logistic Regression



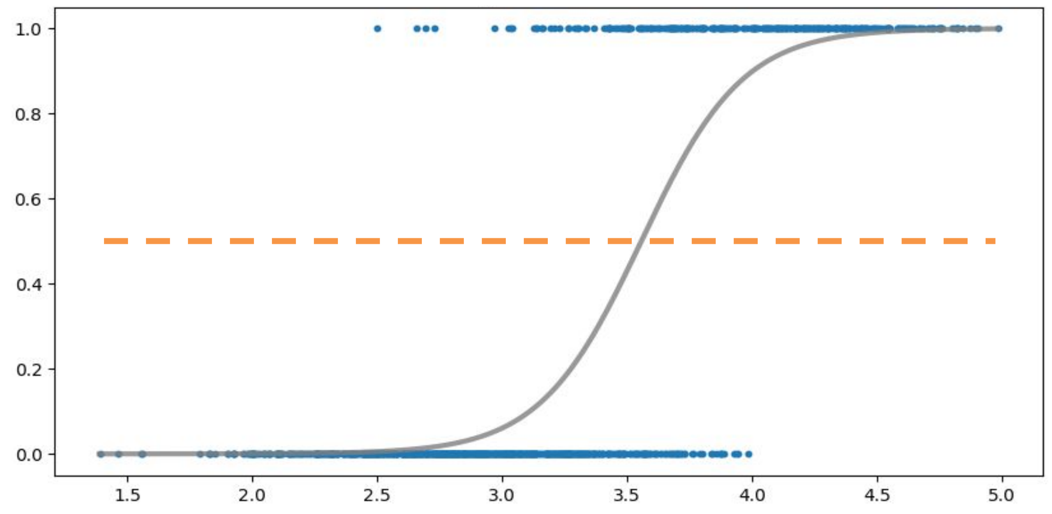
Naive Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Back to Agenda

Estimator

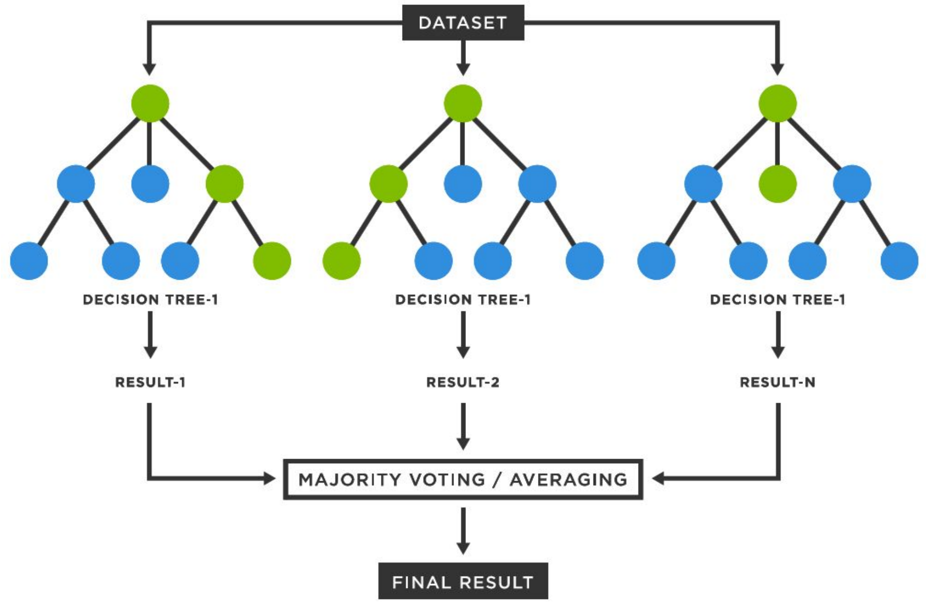
Logistic Regression



Naive Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

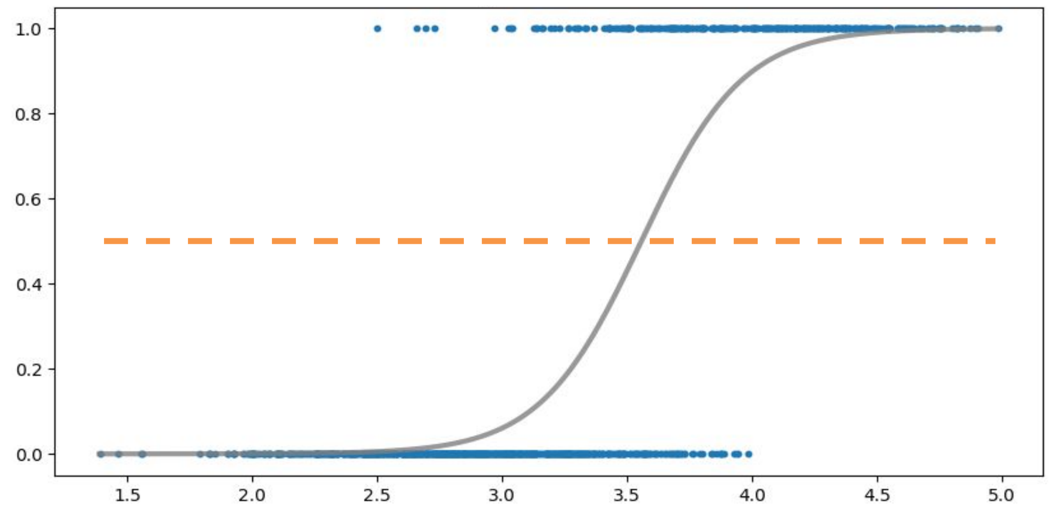
Random Forest



Back to Agenda

Estimator

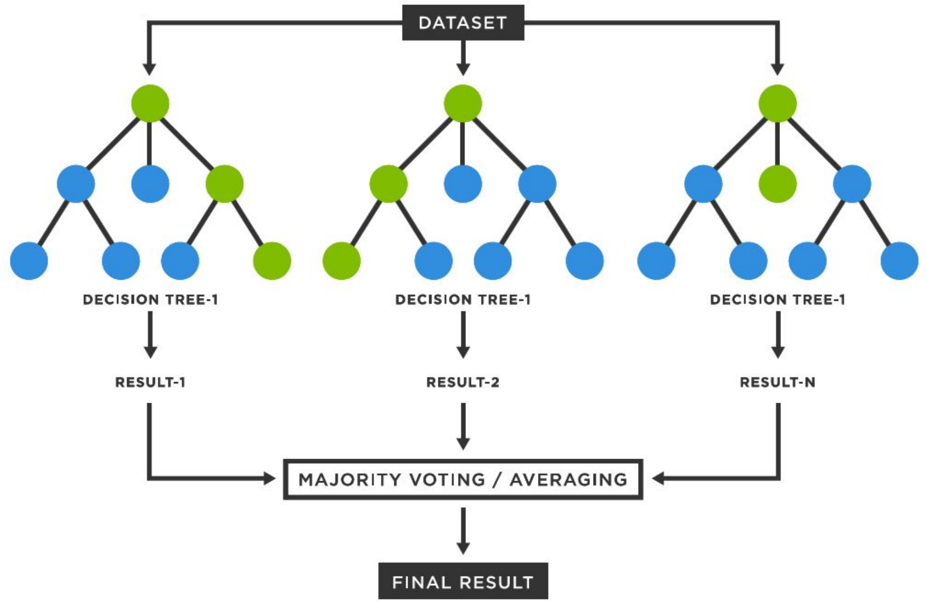
Logistic Regression



Naive Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Random Forest



Good at handling **sparse high-dimensional data**

Back to Agenda

Model - Metrics

Sensitivity - Correctly Identify Omega Comments

Model - Base scores

Transformer - Estimator

Transformer	Estimator	Processing Time	Train Score	Test Score	Sensitivity	Precision	Specificity
Count Vectorizer	Naive Bayes	2.4s	0.8790	0.8349	0.8571	0.8266	0.8118
	Logistic Regression	4.4s	0.9129	0.8266	0.8444	0.8216	0.8080
	Random Forest	7.3s	0.7938	0.7795	0.9468	0.7146	0.6043
TF-IDF Vectorizer	Naive Bayes	2.2s	0.8761	0.8302	0.8355	0.8329	0.8246
	Logistic Regression	3.0s	0.8876	0.8332	0.8844	0.8077	0.7797
	Random Forest	10.1s	0.7893	0.7781	0.9733	0.7050	0.5738

Model - Base scores

Transformer - Estimator

Transformer	Estimator	Processing Time	Train Score	Test Score	Sensitivity	Precision	Specificity
Count Vectorizer	Naive Bayes	2.4s	0.8790	0.8349	0.8571	0.8266	0.8118
	Logistic Regression	4.4s	0.9129	0.8266	0.8444	0.8216	0.8080
	Random Forest	7.3s	0.7938	0.7795	0.9468	0.7146	0.6043
TF-IDF Vectorizer	Naive Bayes	2.2s	0.8761	0.8302	0.8355	0.8329	0.8246
	Logistic Regression	3.0s	0.8876	0.8332	0.8844	0.8077	0.7797
	Random Forest	10.1s	0.7893	0.7781	0.9733	0.7050	0.5738

Model - Base scores

Transformer - Estimator

Transformer	Estimator	Processing Time	Train Score	Test Score	Sensitivity	Precision	Specificity
Count Vectorizer	Naive Bayes	2.4s	0.8790	0.8349	0.8571	0.8266	0.8118
	Logistic Regression	4.4s	0.9129	0.8266	0.8444	0.8216	0.8080
	Random Forest	7.3s	0.7938	0.7795	0.9468	0.7146	0.6043
TF-IDF Vectorizer	Naive Bayes	2.2s	0.8761	0.8302	0.8355	0.8329	0.8246
	Logistic Regression	3.0s	0.8876	0.8332	0.8844	0.8077	0.7797
	Random Forest	10.1s	0.7893	0.7781	0.9733	0.7050	0.5738

Model - Base scores

Transformer - Estimator

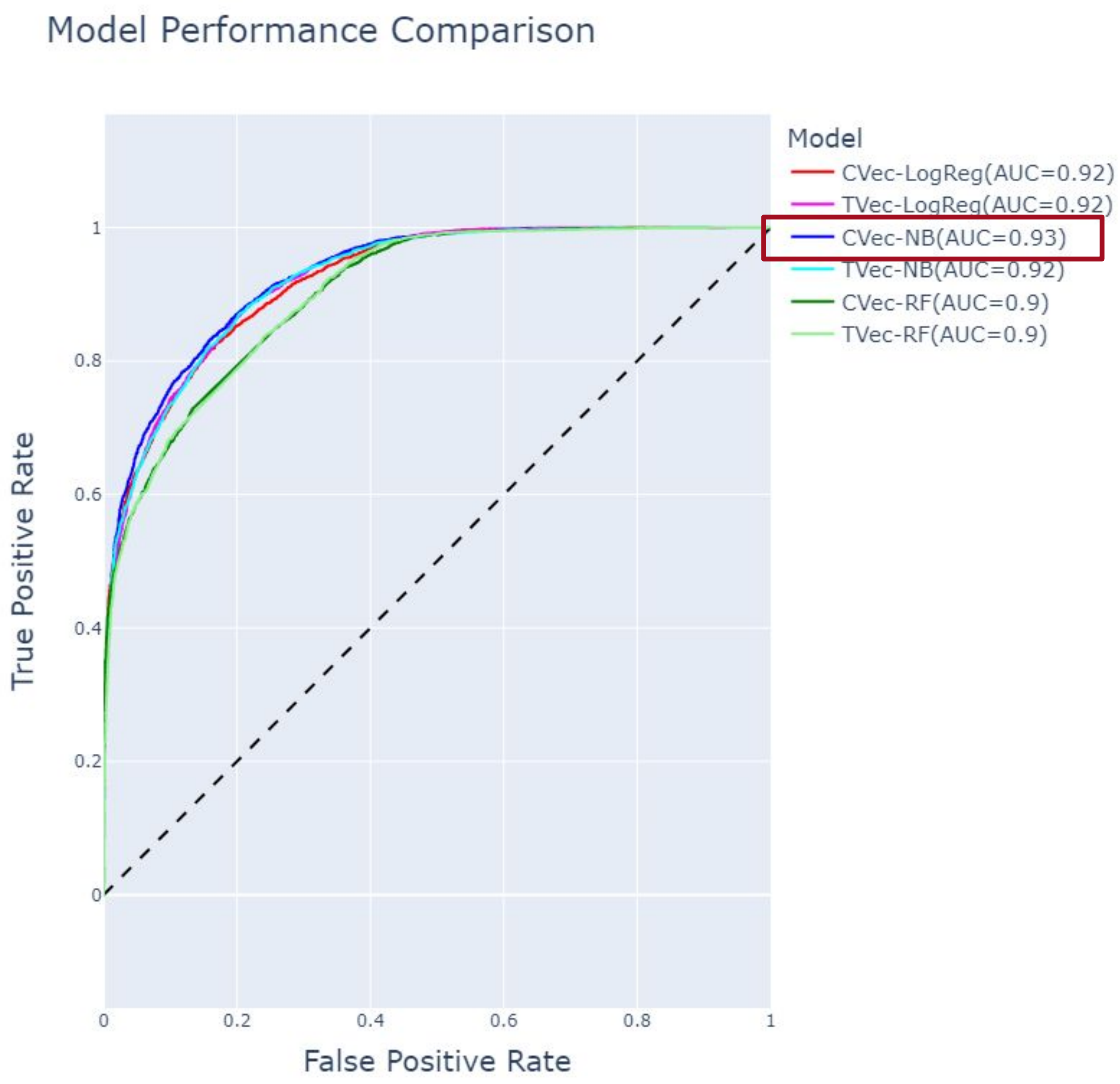
Transformer	Estimator	Processing Time	Train Score	Test Score	Sensitivity	Precision	Specificity
Count Vectorizer	Naive Bayes	2.4s	0.8790	0.8349	0.8571	0.8266	0.8118
	Logistic Regression	4.4s	0.9129	0.8266	0.8444	0.8216	0.8080
	Random Forest	7.3s	0.7938	0.7795	0.9468	0.7146	0.6043
TF-IDF Vectorizer	Naive Bayes	2.2s	0.8761	0.8302	0.8355	0.8329	0.8246
	Logistic Regression	3.0s	0.8876	0.8332	0.8844	0.8077	0.7797
	Random Forest	10.1s	0.7893	0.7781	0.9733	0.7050	0.5738

Model - Base scores

Transformer - Estimator

Transformer	Estimator	Processing Time	Train Score	Test Score	Sensitivity	Precision	Specificity
Count Vectorizer	Naive Bayes	**2.4s**	**0.8790**	**0.8349**	**0.8571**	**0.8266**	**0.8118**
	Logistic Regression	4.4s	0.9129	0.8266	0.8444	0.8216	0.8080
	Random Forest	7.3s	0.7938	0.7795	0.9468	0.7146	0.6043
TF-IDF Vectorizer	Naive Bayes	2.2s	0.8761	0.8302	0.8355	0.8329	0.8246
	Logistic Regression	3.0s	0.8876	0.8332	0.8844	0.8077	0.7797
	Random Forest	10.1s	0.7893	0.7781	0.9733	0.7050	0.5738

Model Evaluation



AUC

Area Under the ROC Curve

- High AUC (Close to 1)
  - Correctly distinguishes between the positive and negative classes
  - Robust and stable in predicting with confidence

Back to Agenda

Model Hypertuning

Transfomer	Estimator	Processing Time	Train Score	Test Score	Sensitivity	Precision	Specificity
Count Vectorizer	Naive Bayes	2.44s	0.8790	0.8349	0.8571	0.8266	0.8118
Count Vectorizer (Tuned)	Naive Bayes (Tuned)	2.16s	0.8642	0.8357	0.8622	0.8246	0.8080

Cvec-NB

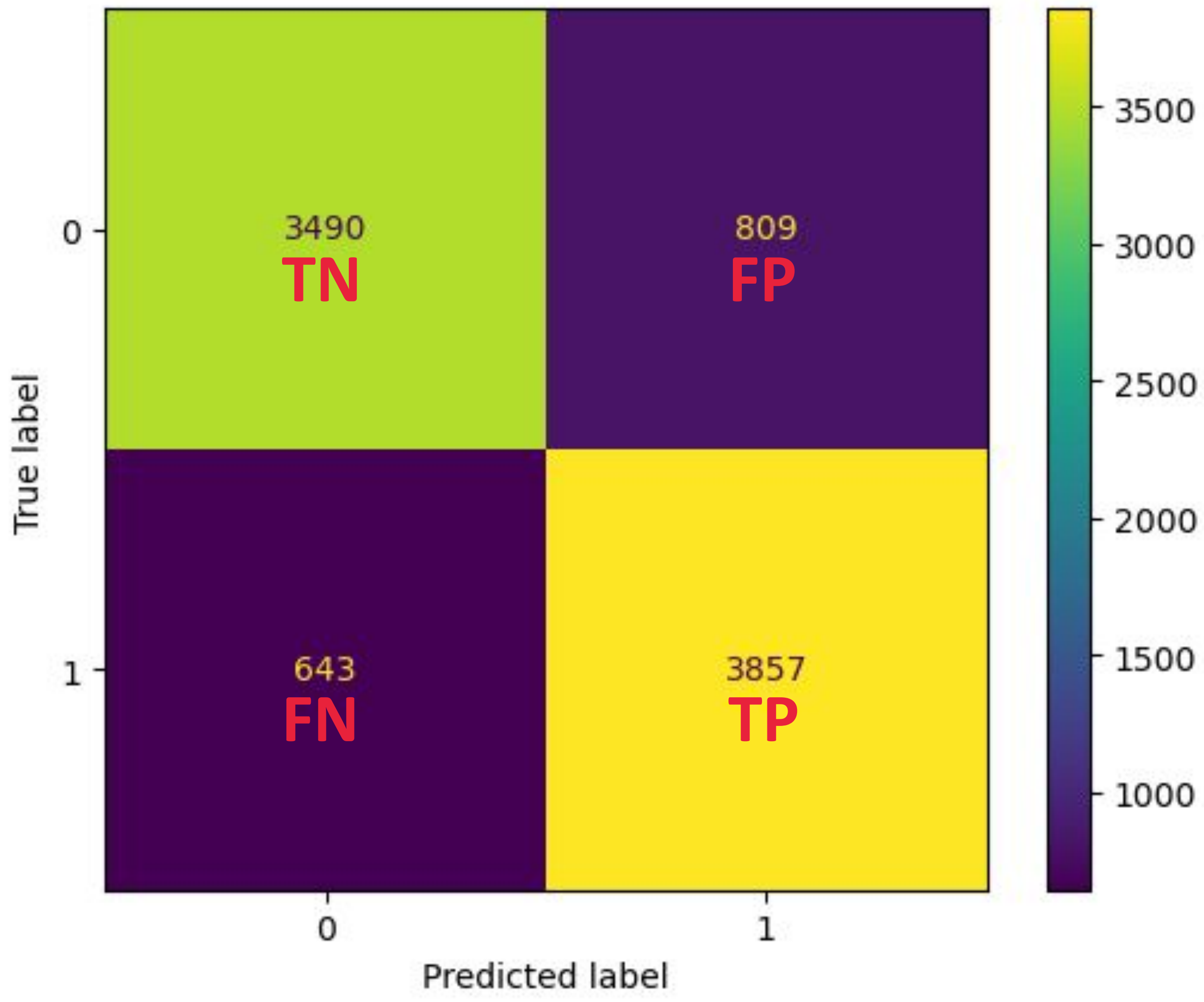
- cvec__max_features = 10,650
- cvec__min_df = 2
- cvec__max_df = 0.85
- cvec__ ngram_rang = (1,1)
- nb__alpha = 0.85

Back to Agenda

Model Evaluation

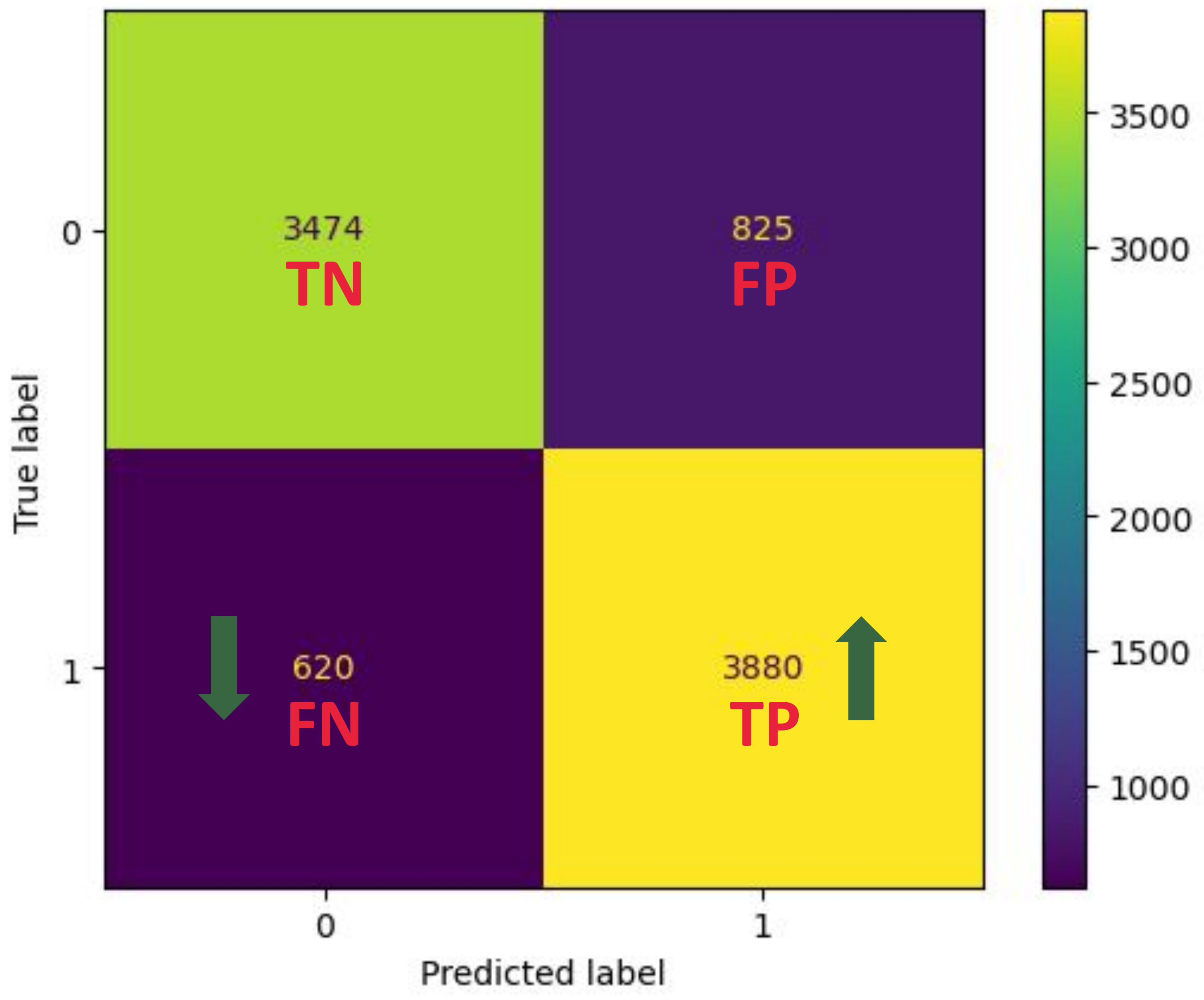
Confusion Matrix

Before Hyperparameter Tuning

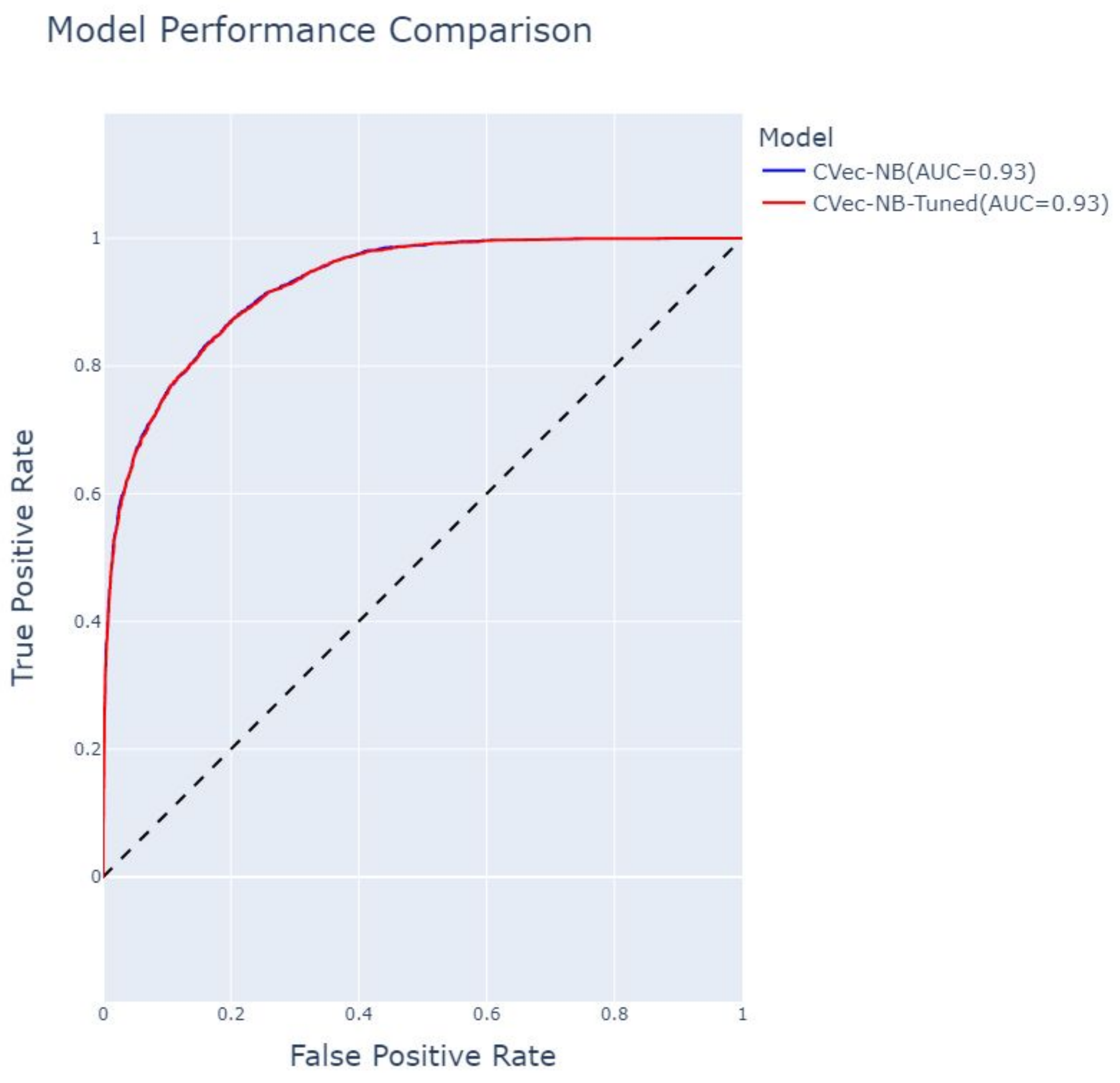


Sensitivity

After Hyperparameter Tuning



Model Evaluation



AUC

Area Under the ROC Curve

- High AUC (Close to 1)
  - correctly distinguishes between the positive and negative classes
  - Robust and stable in predicting with confidence

Model Conclusion

Transfomer	Estimator	Processing Time	Train Score	Test Score	Sensitivity	Precision	Specificity
Count Vectorizer	Naive Bayes	2.44s	0.8790	0.8349	0.8571	0.8266	0.8118
Count Vectorizer (Tuned)	Naive Bayes (Tuned)	**2.16s**	**0.8642**	**0.8357**	**0.8622**	0.8246	0.8080

- Count Vectorizer - Naive Bayes is the best and fastest model
- Higher Sensitivity means better at correctly identifying Omega comments
- Higher Test Score and excellent generalization capability.

Back to Agenda

The Problem Statement

- Data-driven marketing decisions will help us reclaim #2 in industry ranking, it guides organisational objectivity and standard procedures.
- Objectivity remains the plague of our brand and advertising.
- How might we leverage natural language processing to inform our marketing decisions with confidence and objectivity, especially on copy that goes to brand taglines and marketing slogans?

Scan this QR

How well do our slogans score?

OMEGA

Elegance in motion - Constellation.

Timeless sophistication - De Ville

Precise in every detail - Aqua Terra

Cartier

Timeless elegance, defined by Tank

Elevate your elegance with Ballon Bleu

Cartier: Love is All (2022)

How well do our slogans score?

OMEGA

Cartier

Elegance in motion ⟩	39.7% OMEGA	Timeless elegance ⟩	87% Cartier
Timeless sophistication ⟩	13.6% OMEGA	Elevate your elegance ⟩	75.8% Cartier
Precise in every detail ⟩	86.9% OMEGA	Love is All ⟩	30% Cartier

Can we make a confident decision here?

Which word sounds more OMEGA?

“Be a master of _____”

Precision?
Sophistication?
Elegance?
Exploration?

Conclusion and Recommendation

OBJECTIVITY

Measure how well an ad copy will resonate with your audience

DIRECTION

Amplify and solidify use of words that are shared languages with customer

CONFIDENCE

Additional layer of validation to give confidence in publishing an ad

TIMELINESS

Reduces time taken to assess and evaluate effectiveness of copy

SOURCE OF TRUTH

Shared source of truth between client and agency

Limitations and Future enhancements

Limitations

- This classifier only takes into account text from customers between OMEGA and Cartier. While our main competitor is Cartier, we should also take into account other watch brands.
- The model only takes words that were used in comments, however does not take into account words with similar meanings e.g. precision vs accuracy.
- The current model does not take into account sentiment scores, which could help to identify feedback for future campaigns.
- In its purest form, this model could become a clutch that restricts creativity. Brands and customers also evolve to speak in a way that is never before done.

Future enhancements

- For the future, we will include forum posts from the top 5 watch brands to create a multi-classification model that provides a more robust view of the industry.
- We can explore other transformer libraries such as BERT that is more robust in semantics, contextual meanings and relationship between similar words.
- A quick addition to the model could show sentiment comparisons between top 5 watch brands.
- The typical process taken to develop a creative ad, slogan and tagline should still be common practice as it’s designed with much intention. The model serves only as a layer of validation.

Confusion Matrix

	Predicted Positive	Predicted Negative	
Actual Positive	TP *True Positive*	FN *False Negative*	Sensitivity $\frac{TP}{(TP + FN)}$
Actual Negative	FP *False Positive*	TN *True Negative*	Specificity $\frac{TN}{(TN + FP)}$
	Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Our Aim

To propel Omega to #2 ranking by **strengthening Omega's brand community**

Through means of Natural Language Processing, we created a text-based classification model that aims to distinct Omega's branding and advertising efforts from Cartier in the following ways:

#1

Systemize the brand's **voice and tone**

Are we communicating in a way our loyalists can relate to?

#2

Leverage well-loved **features**

Are we able to identify distinct features that our loyalists love?

#3

Identify new **channels**

Outside of existing communities, where else can we engage potential customers?

Model Evaluation

Model	Processing Time	Train Score	Test Score
CVec-LR	32.5s	0.8980	0.8261
TVec-LR	36.1s	0.9203	0.8335
CVec-NB	17.9s	08642	0.8357
TVec-NB	17.8s	0.8778	0.8315

Back to Agenda

Model Evaluation

Model	Processing Time	Train Score	Test Score
CVec-LR	32.5s	**0.8980** Overfitting	**0.8261**
TVec-LR	36.1s	**0.9203** Overfitting	**0.8335**
CVec-NB	17.9s	08642	0.8357
TVec-NB	17.8s	0.8778	0.8315

Back to Agenda

Model Evaluation

Best Parameters

Cvec-NB

- cvec__max_features = 10,650
- cvec__min_df = 2
- cvec__max_df = 0.85
- cvec__ ngram_rang = (1,1)
- nb__alpha = 0.85

Tvec-NB

- cvec__max_features = 10,000
- cvec__min_df = 2
- cvec__max_df = 0.85
- cvec__ ngram_rang = (1,1)
- nb__alpha = 0.7

Model Evaluation

Model	Processing Time	Train Score	Test Score	Sensitivity	Precision
CVec-NB	17.5s	08642	0.8357	0.8622	0.8246
Tvec-NB	17.9s	0.8778	0.8315	0.8364	0.8345

Back to Agenda

Model Evaluation

Model	Processing Time	Train Score	Test Score	Sensitivity	Precision
CVec-NB	17.5s	0.8642	0.8357	**0.8622**	0.8246
Tvec-NB	17.9s	0.8778	0.8315	0.8364	0.8345

We decided to choose **Count Vectorizer - Naive Bayes**, because we want to correctly identify Omega comments (**higher Sensitivity**).

Back to Agenda

Model

Time and Resource Management

- Run different models in different machines
- Run hyperparameters tuning in steps

Model

Hyperparameters Tuning Elimination

cvec__max_features = 5,000, 7,500, 10,000

cvec__min_df = 2, 3

cvec__max_df = 0.7, 0.8, 0.9

cvec__ngram_range = (1, 1), (1, 2)

lr__C = 0.1, 1, 10

lr__penalty = L1, L2

lr__solver = liblinear, lbfgs

Back to Agenda

Model

Hyperparameters Tuning Elimination

cvec__max_features = 5,000, 7,500, **10,000**

cvec__min_df = **2**, 3

cvec__max_df = 0.7, **0.8**, 0.9

cvec__ngram_range = **(1, 1)**, (1, 2)

lr__C = **0.1**, 1, 10

lr__penalty = L1, **L2**

lr__solver = liblinear, **lbfgs**

Model

Hyperparameters Tuning Elimination

cvec__max_features = 5,000, 7,500, **10,000**

cvec__min_df = **2**, 3

cvec__max_df = 0.7, **0.8**, 0.9

cvec__ngram_range = **(1, 1)**, (1, 2)

lr__C = **0.1**, 1, 10

lr__penalty = L1, **L2**

lr__solver = liblinear, **lbfgs**

→

cvec__max_features = 9,000, 11,000, 13,000

cvec__min_df = **2**

cvec__max_df = **0.8**

cvec__ngram_range = **(1, 1)**

lr__C = 0.01, 0.2, 0.4

lr__penalty = **L2**

lr__solver = **lbfgs**

Model

Hyperparameters Tuning Elimination

cvec__max_features = 5,000, 7,500, **10,000**

cvec__min_df = **2**, 3

cvec__max_df = 0.7, **0.8**, 0.9

cvec__ngram_range = **(1, 1)**, (1, 2)

lr__C = **0.1**, 1, 10

lr__penalty = L1, **L2**

lr__solver = liblinear, **lbfgs**

→

cvec__max_features = 9,000, **11,000**, 13,000

cvec__min_df = **2**

cvec__max_df = **0.8**

cvec__ngram_range = **(1, 1)**

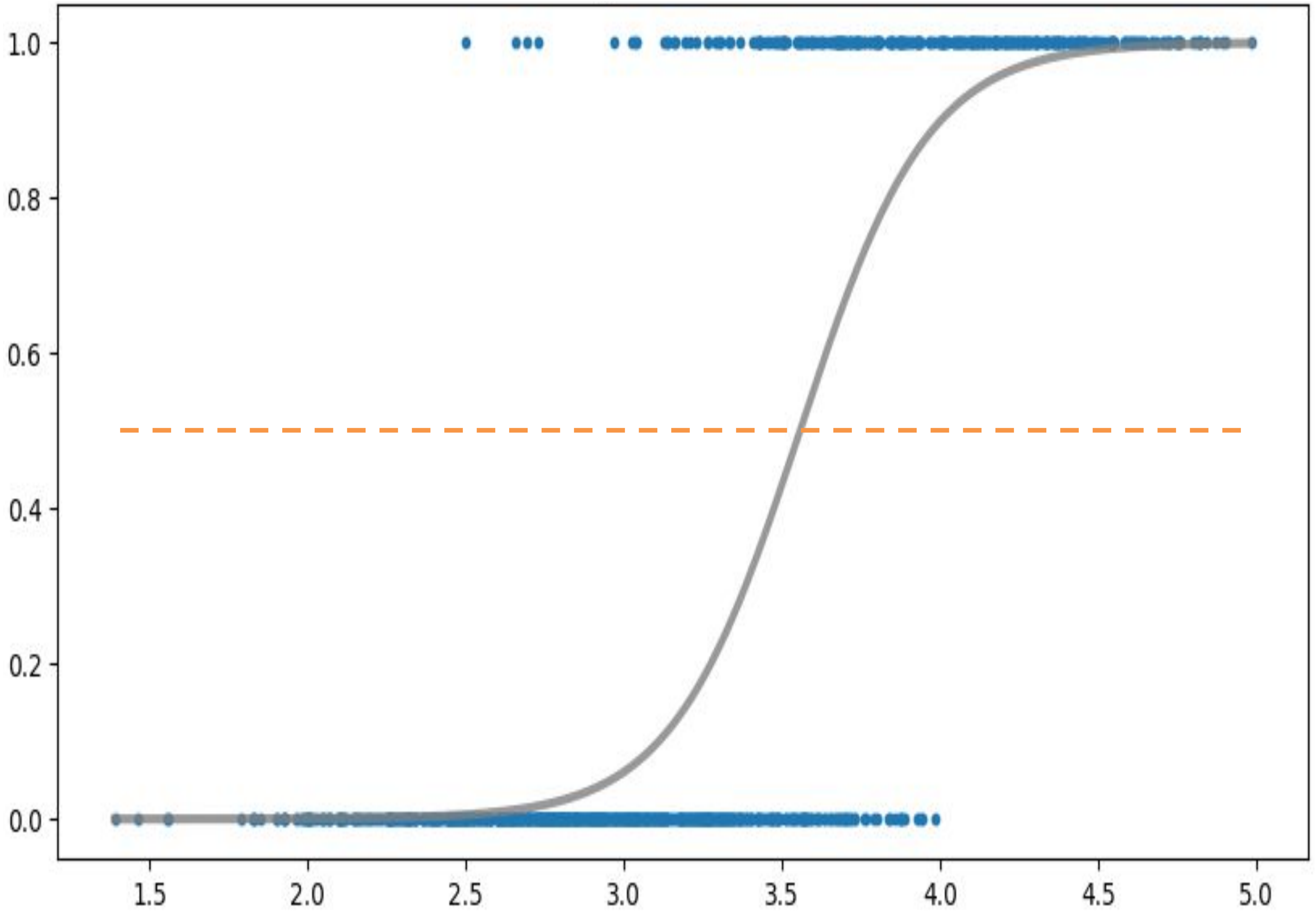
lr__C = 0.01, 0.2, **0.4**

lr__penalty = **L2**

lr__solver = **lbfgs**

Estimator

Logistic Regression

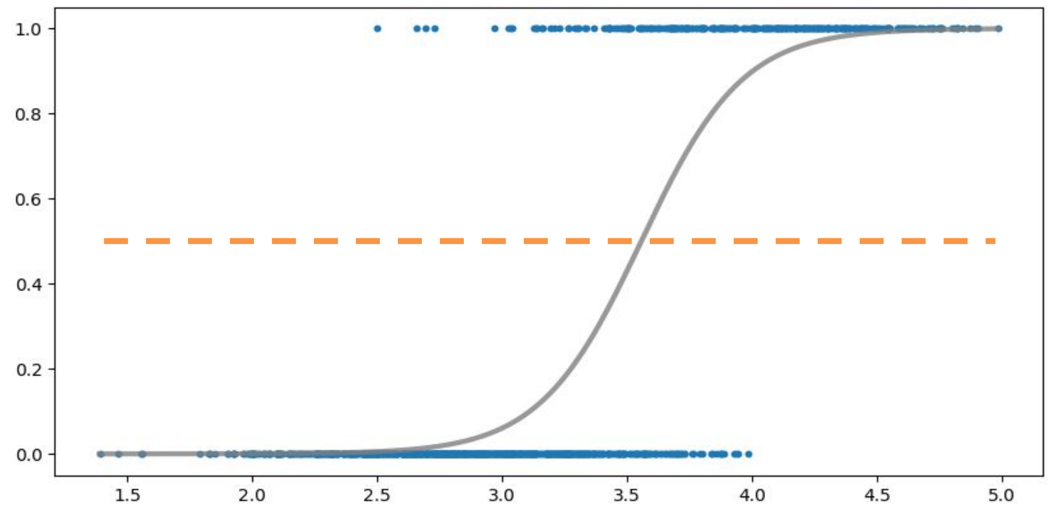


Why Logistic Regression:

- Logistic regression is a simple and interpretable model that works well when the relationship between features and target is approximately linear.
- It can handle sparse high-dimensional data, which is common in text classification.

Estimator

Logistic Regression

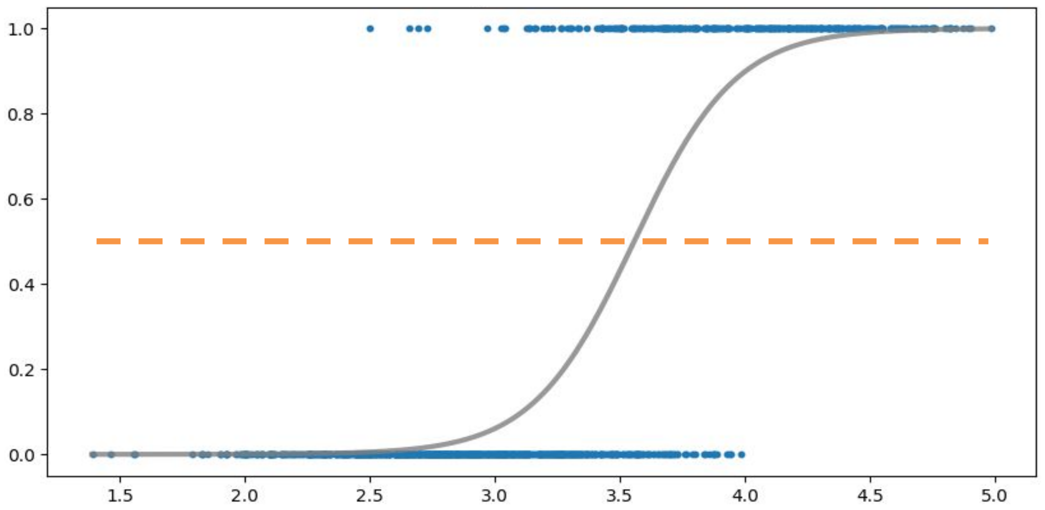


- Simple
- Easy to estimate when relationships are linear
- Sparse, high dimensional data (e.g. text)

Back to Agenda

Estimator

Logistic Regression



- Simple
- Easy to estimate when relationships are linear
- Sparse, high dimensional data (e.g. text)

Naive Bayes

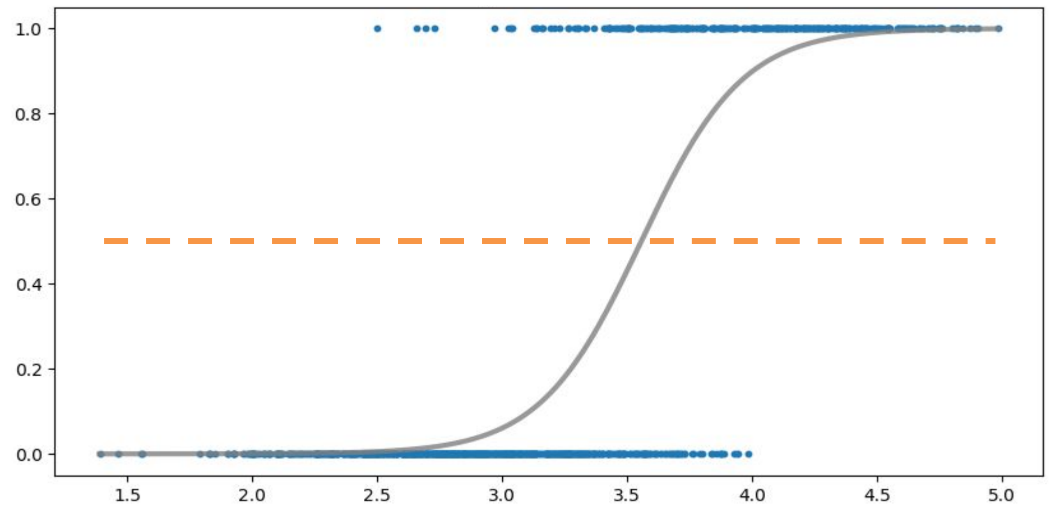
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Efficient for high-dimensional data
- Assumes features are independent
- Robust to irrelevant feature
- Handles noisy data

Back to Agenda

Estimator

Logistic Regression



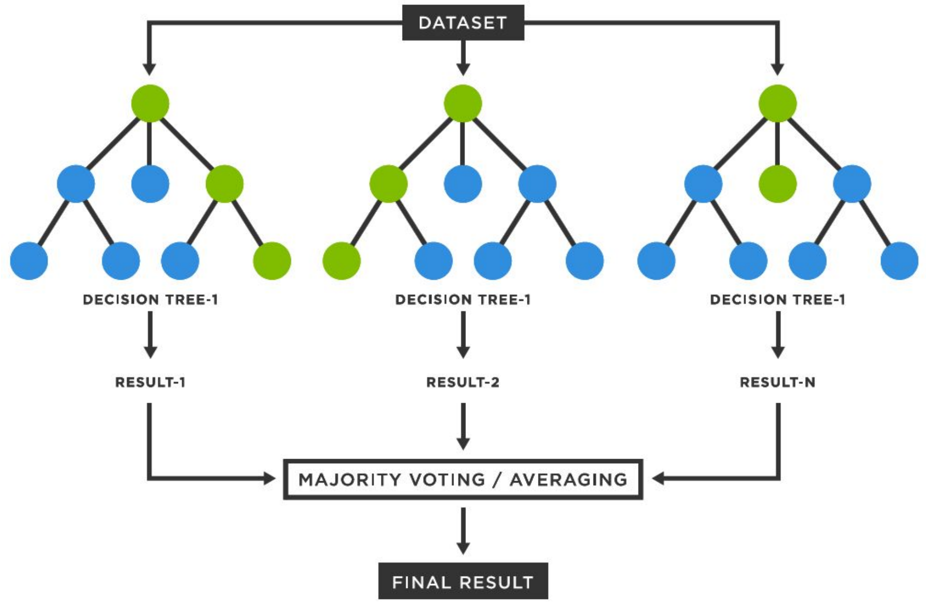
- Simple
- Easy to estimate when relationships are linear
- Efficient for high-dimensional data

Naive Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Efficient for high-dimensional data
- Assumes features are independent
- Robust to irrelevant feature
- Handles noisy data

Random Forest



- Can Capture non-linear features relationships
- Efficient for high-dimensional data and less overfitting
- Robust to irrelevant feature

Back to Agenda

Model - Base scores

Transformer - Estimator

Transfomer	Estimator	Processing Time	Train Score	Test Score	Sensitivity	Precision	Specificity
Count Vectorizer	Naive Bayes	2.5s	0.8790	0.8349	**0.8571**	0.8266	0.8118
	Logistic Regression	**5.5s**	0.9129	0.8266	**0.8444**	0.8216	0.8080
TF-IDF Vectorizer	Naive Bayes	2.5s	0.8761	0.8302	**0.8355**	0.8329	0.8246

Optimising for Sensitivity

- Focused on correctly predicting OMEGA
- Reduce OMEGA comments classified as Cartier

Model - Base scores

Transformer - Estimator

Transfomer	Estimator	Processing Time	Train Score	Test Score	Sensitivity	Precision	Specificity
Count Vectorizer	Naive Bayes	2.5s	0.8790	0.8349	**0.8571**	0.8266	0.8118
	Logistic Regression	~~5.5s~~	~~0.9129~~	~~0.8266~~	~~0.8444~~	~~0.8216~~	~~0.8080~~
TF-IDF Vectorizer	Naive Bayes	~~2.5s~~	~~0.8761~~	~~0.8302~~	~~0.8355~~	~~0.8329~~	~~0.8246~~

Optimising for Sensitivity

- Focused on correctly predicting OMEGA
- Reduce OMEGA comments classified as Cartier

What is the importance of brand voice?

Brand voice matters for several reasons. Here are a few of them.

1. Brand voice differentiates your organization

Your brand voice is a big part of what sets you apart from the rest. This is particularly true in today's overcrowded online environment. You'd be surprised how little time you have to make a lasting, positive impression on your target audience.

It's not what you say, it's *how* you say it that attracts attention and strengthens your credibility. Because, after all, *what* you say may not be unique.

A similar company or organization may have a similar message. Yet, if the audiences you're competing for jive more with how you present that message, you can gain the advantage. (The opposite is also true, which is why it's important to invest in developing a voice that will stand the test of time).

The Importance Of Brand Voice And Tone

Forbes Agency Council

Jonathan Forrester For
Forbes Agency Council

How can it help gain customers?

Voice and tone collaborate with each other to strengthen your brand's appeal to consumers. By evaluating the responses received and encouraging ongoing conversation, you can identify the type of audience you're actively engaging with. The way content is delivered to readers is determined by the voice, tone and language. With authentic connections and increasing organic traffic about your brand, your company's messages are more likely to be welcomed and happily received.

Customers are only able to communicate openly when they are listened to and in return, may bring in other prospective clients. You can improve the voice and tone you employ by listening to the way your community describes your brand and shares their experiences.

Source