User Aesthetics Identification for Fashion Recommendations

Liwei Liu, Ivo Silva, Pedro Nogueira, Ana Magalhães, **Eder Martins**

FARFETCH

AGENDA

1. Aesthetics
2. Models & Features
3. Experiments
4. Conclusions & Final Remarks

AESTHETICS

Challenges on fashion recommendation

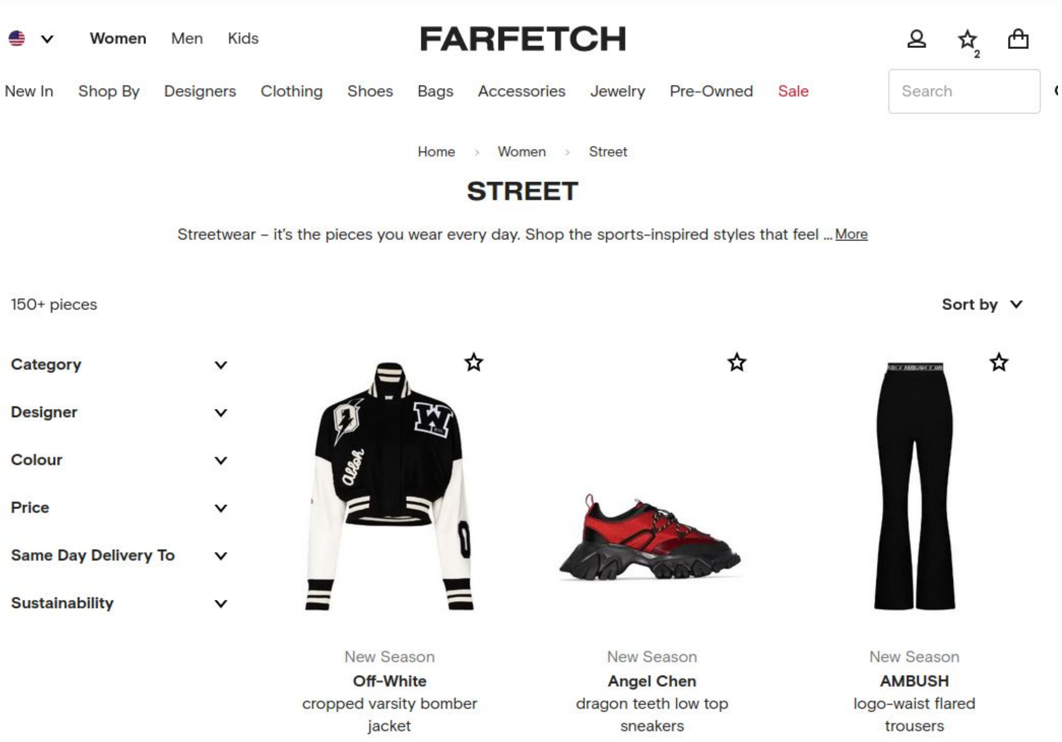
- Unique sense of style
- Personalize the experience
- Incorporate the concepts of fashion and style in our recommenders
- Inspire our customers

Aesthetic concepts

- Defined by our fashion experts
  - Reflect our customer's style and preferences
- Having a dedicated listing page
  - Curated list of products

Aesthetic concepts

- Female
  - Arty
  - Classic
  - Edgy
  - Feminine
  - Minimal
  - Streetwear



Aesthetic concepts

- Expert domain knowledge
- User behavior

Is it possible to predict the aesthetics of our customers?

Models & Features

A customer could be interested in more than one aesthetics

Multi label classification problem

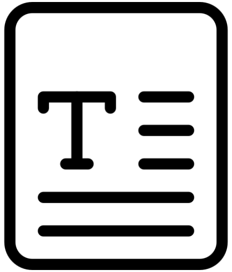
Models

- Random Forest
  - Binary relevance
  - Label powerset
- CNN

Features

Statistics

Images



Texts

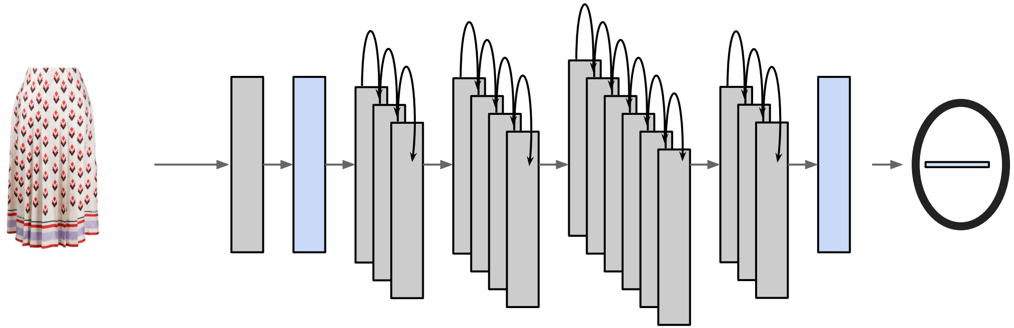
User statistics

- Number sessions
- Number orders
- ...
- Customer region
- Device
- Categories a user has interacted with
- Brands a user has interacted with

Weighted sum of user actions on those categories and brands

Image Features

ResNet 50

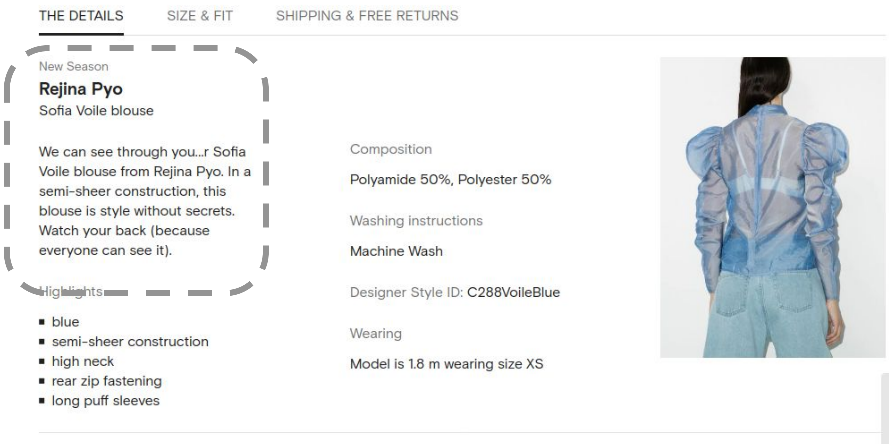


Final layer (1x2048 vector)

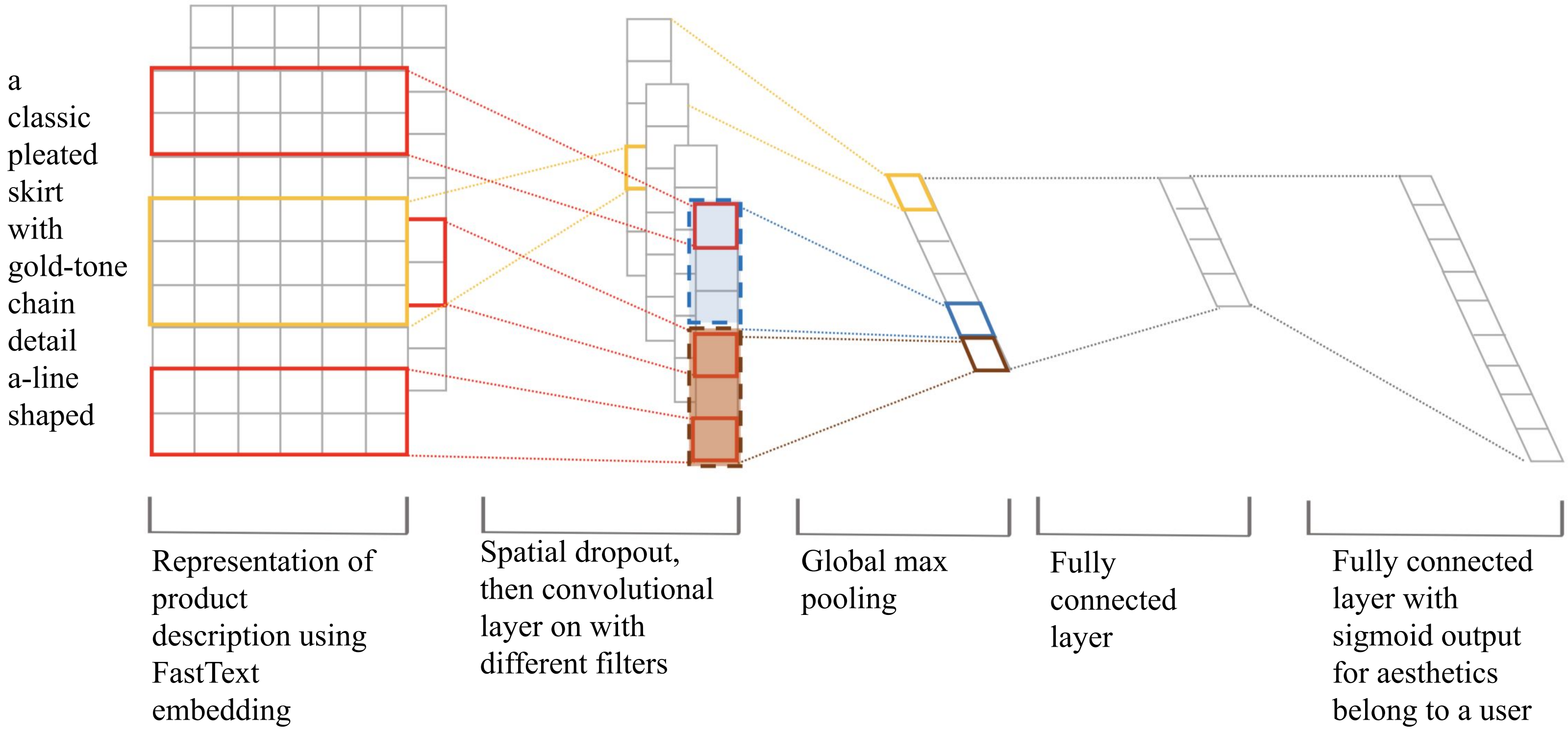
- Avg / min / max their interacted product embeddings
- Cluster all the products embeddings per category

Textual features

- Tokens/words of all the products a user has interacted with in the training data as features
- TF-IDF
- Word embedding - FastText
  - 1x300 vector

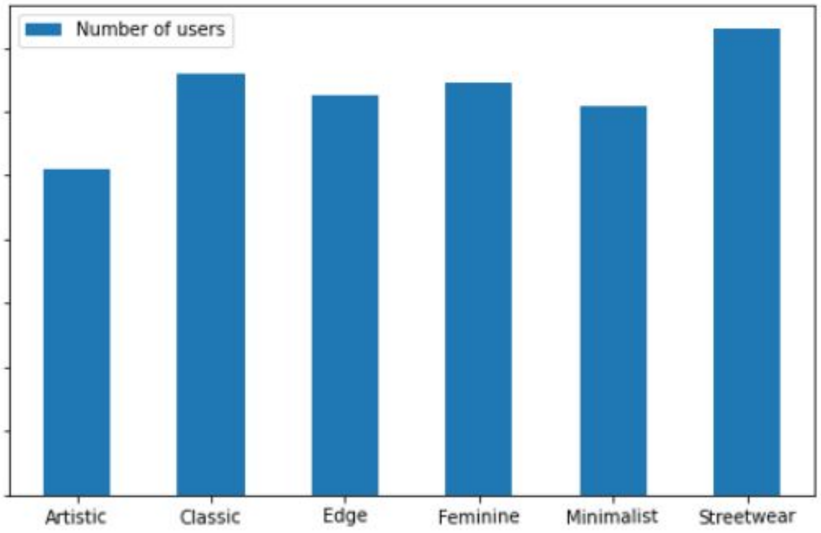


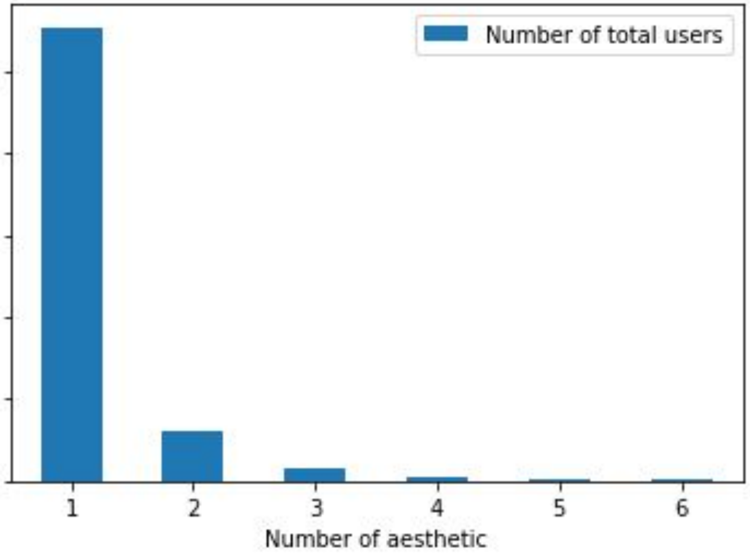
CNN with FastText embeddings



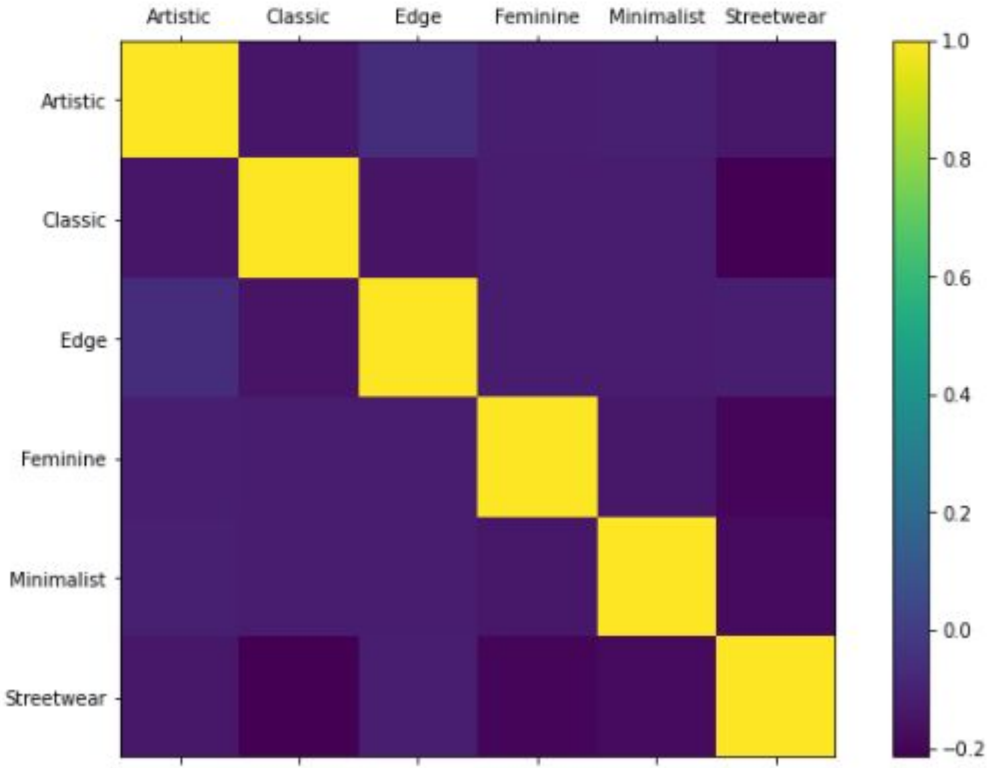
Experiments

Fairly balanced





Aesthetics are not correlated with each other



Binary Random Forest classification models

Features	F1	Precision	Recall
Word embedding (TF-IDF)	0.525	0.586	0.476
Word embedding (TF-IDF) + Image embedding	0.524	0.585	0.474
General Stats + Word embedding (TF-IDF) + Image embedding	0.507	0.513	0.500
General Stats + Word embedding (TF-IDF)	0.503	0.521	0.486
General Stats	0.442	0.360	0.572
General Stats + Image embedding	0.420	0.382	0.467
Word embeddings (FastText)	0.418	0.335	0.554
Image embedding (clusters)	0.348	0.295	0.424
Random	0.257	0.197	0.370

Binary Random Forest classification models

Single features

Features	F1	Precision	Recall
Word embedding (TF-IDF)	**0.525**	**0.586**	**0.476**
Word embedding (TF-IDF) + Image embedding	0.524	0.585	0.474
General Stats + Word embedding (TF-IDF) + Image embedding	0.507	0.513	0.500
General Stats + Word embedding (TF-IDF)	0.503	0.521	0.486
General Stats	**0.442**	**0.360**	**0.572**
General Stats + Image embedding	0.420	0.382	0.467
Word embeddings (FastText)	**0.418**	**0.335**	**0.554**
Image embedding (clusters)	**0.348**	**0.295**	**0.424**
Random	0.257	0.197	0.370

Binary Random Forest classification models

All feature combinations

Features	F1	Precision	Recall
Word embedding (TF-IDF)	0.525	0.586	0.476
Word embedding (TF-IDF) + Image embedding	**0.524**	**0.585**	**0.474**
General Stats + Word embedding (TF-IDF) + Image embedding	**0.507**	**0.513**	**0.500**
General Stats + Word embedding (TF-IDF)	**0.503**	**0.521**	**0.486**
General Stats	0.442	0.360	0.572
General Stats + Image embedding	**0.420**	**0.382**	**0.467**
Word embeddings (FastText)	0.418	0.335	0.554
Image embedding (clusters)	0.348	0.295	0.424
Random	0.257	0.197	0.370

Binary Random Forest classification models

Random Forest with Word embedding (TF-IDF) features performs better.

Some words, in product descriptions could be used as a strong indicator of aesthetics preferences.

Features	F1	Precision	Recall
Word embedding (TF-IDF)	**0.525**	**0.586**	**0.476**
Word embedding (TF-IDF) + Image embedding	0.524	0.585	0.474
General Stats + Word embedding (TF-IDF) + Image embedding	0.507	0.513	0.500
General Stats + Word embedding (TF-IDF)	0.503	0.521	0.486
General Stats	0.442	0.360	0.572
General Stats + Image embedding	0.420	0.382	0.467
Word embeddings (FastText)	0.418	0.335	0.554
Image embedding (clusters)	0.348	0.295	0.424
Random	0.257	0.197	0.370

Binary Random Forest classification models

Image embedding perform the worst.

In the same aesthetic products could look very different, resulting in very different embeddings, which in turn confuses Random Forest.

Features	F1	Precision	Recall
Word embedding (TF-IDF)	0.525	0.586	0.476
Word embedding (TF-IDF) + Image embedding	0.524	0.585	0.474
General Stats + Word embedding (TF-IDF) + Image embedding	0.507	0.513	0.500
General Stats + Word embedding (TF-IDF)	0.503	0.521	0.486
General Stats	0.442	0.360	0.572
General Stats + Image embedding	0.420	0.382	0.467
Word embeddings (FastText)	0.418	0.335	0.554
Image embedding (clusters)	**0.348**	**0.295**	**0.424**
Random	0.257	0.197	0.370

Random Forest vs CNN

Time range	Algorithm	Features	F1	Precision	Recall
6 months	RF	Word embedding (TF-IDF)	0.525	0.586	0.476
3 months	RF	Word embedding (TF-IDF)	0.505	0.555	0.463
6 months	CNN	Word embedding (FastText)	0.404	0.680	0.288
3 months	CNN	Word embedding (FastText)	0.307	0.687	0.199

Random Forest vs CNN

The addition of data have little impact on the model quality

Time range	Algorithm	Features	F1	Precision	Recall
6 months	**RF**	**Word embedding (TF-IDF)**	**0.525**	**0.586**	**0.476**
3 months	**RF**	**Word embedding (TF-IDF)**	**0.505**	**0.555**	**0.463**
6 months	CNN	Word embedding (FastText)	0.404	0.680	0.288
3 months	CNN	Word embedding (FastText)	0.307	0.687	0.199

Random Forest vs CNN

CNN model have a significant increase in recall.

Time range	Algorithm	Features	F1	Precision	Recall
6 months	RF	Word embedding (TF-IDF)	0.525	0.586	0.476
3 months	RF	Word embedding (TF-IDF)	0.505	0.555	0.463
6 months	**CNN**	**Word embedding (FastText)**	**0.404**	**0.680**	**0.288**
3 months	**CNN**	**Word embedding (FastText)**	**0.307**	**0.687**	**0.199**

Random Forest vs CNN

CNN models performed better on precision at the cost of recall leading to a worst F1 when compared with Random Forest.

Time range	Algorithm	Features	F1	Precision	Recall
6 months	RF	Word embedding (TF-IDF)	0.525	0.586	**0.476**
3 months	RF	Word embedding (TF-IDF)	0.505	0.555	**0.463**
6 months	CNN	Word embedding (FastText)	0.404	**0.680**	0.288
3 months	CNN	Word embedding (FastText)	0.307	**0.687**	0.199

Aesthetics results breakdown

Note the positive correlation between class frequency and the evaluation metrics which might be an indicator that more popular aesthetics are easier to classify.

Aesthetic	Frequency	F1	Precision	Recall
Streetwear	**0.230**	**0.635**	0.739	0.557
Classic	**0.207**	**0.589**	0.671	0.524
Feminine	**0.203**	**0.483**	0.554	0.428
Edge	**0.196**	**0.484**	0.551	0.431
Minimalist	**0.192**	**0.499**	0.583	0.437
Artistic	**0.158**	**0.442**	0.525	0.383

Feminine Aesthetic customer example

cowl neck tank mini dress
Maisie Wilen

puff sleeve midi dress
CECILIE BAHNSEN

Livarno satin midi skirt
GAUGE81

June v-neck camisole
BERNADETTE

Kim high neck top
ROTATE

Final Remarks

- Aesthetics could be inferred from our customer shopping behavior
- Future work
  - Test in live
  - Conduct a survey study to improve and validate our models