Farfetch: Size & Fit Assistant

PURDUE UNIVERSITY | The Data Mine

Rashmi Ananth, Seung Hyun Choi, Evzenie Coupkova, Rohan Das, Zhiying Dong, Kanishk Kamalvanshi, Shivani Kogta, Vamsi Kolluri, Seoyoung Lee, Khushi Saini, Tanvi Sattar, Dhruv Sujatha, Derek Sun, Nitesh Wagh

Introduction, Research Problem, Motivation & Goals

- **Introduction:** Farfetch is a British-Portuguese online luxury fashion store, founded in 2007, which has partnered with over 700 boutiques to sell a variety of products
- **Research Problem:** Enhance customer satisfaction through aspects of comfort, accuracy, and fulfillment with the products
- **Motivation:** Reduce the economic costs invested entailed in manufacturing and return rates
- **Overall Goal:** For a given user or product, develop a model whose output can be used to recommend the product with the right size and create a web app which is model agnostic, to explore recommendations based on the user.

Data & Feature Extraction

- **Numerical variables:** converted height, weight into doubles, bust size modification
- **Categorical Variables:** used neural network embedding layer to reduce the dimensionality of the categorical features
- **Additional latent features:** used skip-gram based Word2Vec model to extract latent features from the user purchase history
- **Review text:** NLP methods, sentiment analysis on review relative to size - returns positive, negative, neutral, and compound score

Mapping	Value
Z	-999
a	1
aa	2
b	3
c	4
d	5
d+	6
dd	7
ddd/e	8
f	9
g	10
h	11
i	12
j	13

Modified_review	review_text
An adorable romper Belt zipper little hard nav...	An adorable romper! Belt and zipper were a lit...

Score_Neutral	Score_Negative	Score_Positive	Score_Compound
0.554	0.036	0.410	0.9097

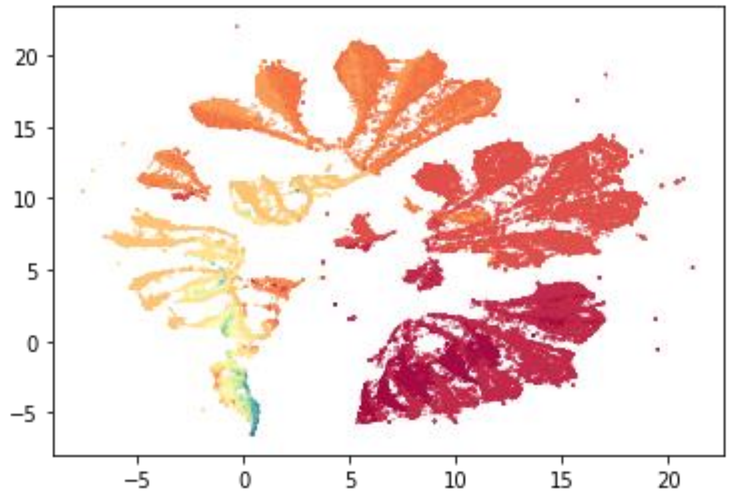

FIGURE 1: Visualization of the structure of our dataset enriched with latent features after dimensionality reduction into 2D space. Different colors correspond to different sizes of products.

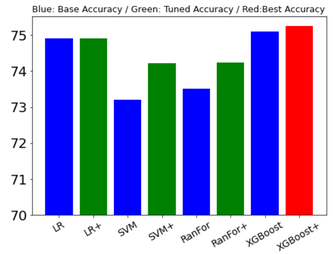

FIGURE 2: Visualization depicting and comparing the performance between the various models.

Blue: Test accuracy without hyperparameter tuning
Green: Test accuracy with hyperparameter tuning
Red: Best accuracy: **75.29%**

Web Application: Fit Assistant

Goal: Explore and preview user data, to present to internal end-users at Farfetch

INPUT	OUTPUT
User ID Product ID	Product ID Predicted size for the product

Backend

MySQL database | Model (via Pickle File)

Database tables

User Information
Product History | Model
Output

Models & Results (Size and Fit prediction)

Benchmarks: Random Prediction, Average Purchase History & Linear Regression

Average Purchase History

- Goal: Predict the size of a user using their past purchase history, and their average sizes
- Outcome:
  - Mean Squared Error (MSE) of 6.13
  - 60% users only have one purchase, which skewed data
  - Improved MSE: 27.67

Random models	Accuracy
Wide Wigner	21.70%
Narrow Wigner	29.77%
Average – Zero Wigner	29.78%
Gaussian	30.51%

Model Type (Final Models)	Test data share	Training accuracy	Testing accuracy
Logistic Regression (without hyperparameter tuning)	33%	74.8590%	74.9058%
Logistic Regression (with hyperparameter tuning)	33%	74.8553%	74.9077%
Support Vector Machines: Linear (without hyperparameter tuning)	33%	76.6000%	73.2025%
Support Vector Machines (with hyperparameter tuning)	33%	75.8000%	74.2144%
Random Forest (without hyperparameter tuning)	33%	99.7739%	73.5173%
Random Forest (with hyperparameter tuning)	33%	74.4437%	74.2444%
XGBoost (without hyperparameter tuning)	33%	77.6611%	75.1045%
XGBoost (with hyperparameter tuning)	33%	76.1465%	75.2881%

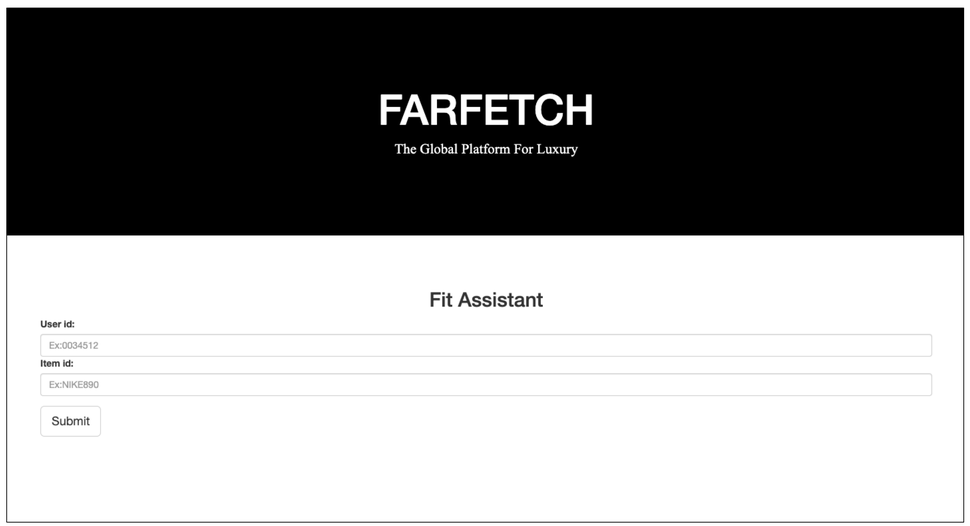

FIGURE 3: Our input webpage for the Farfetch Fit Assistant

Acknowledgements: José Marcelino and Farfetch Corporate Partner mentors, Data Mine staff