FASHION IMAGE CAPTIONING (Image/Missing Data)

In collaboration with FARFETCH

L. Yao, R. Miao, C. Boumansour, M. Zhu, E. Zhao, X. Liu, S. Ahluwalia, T. Dao, M. Dhanushkodi

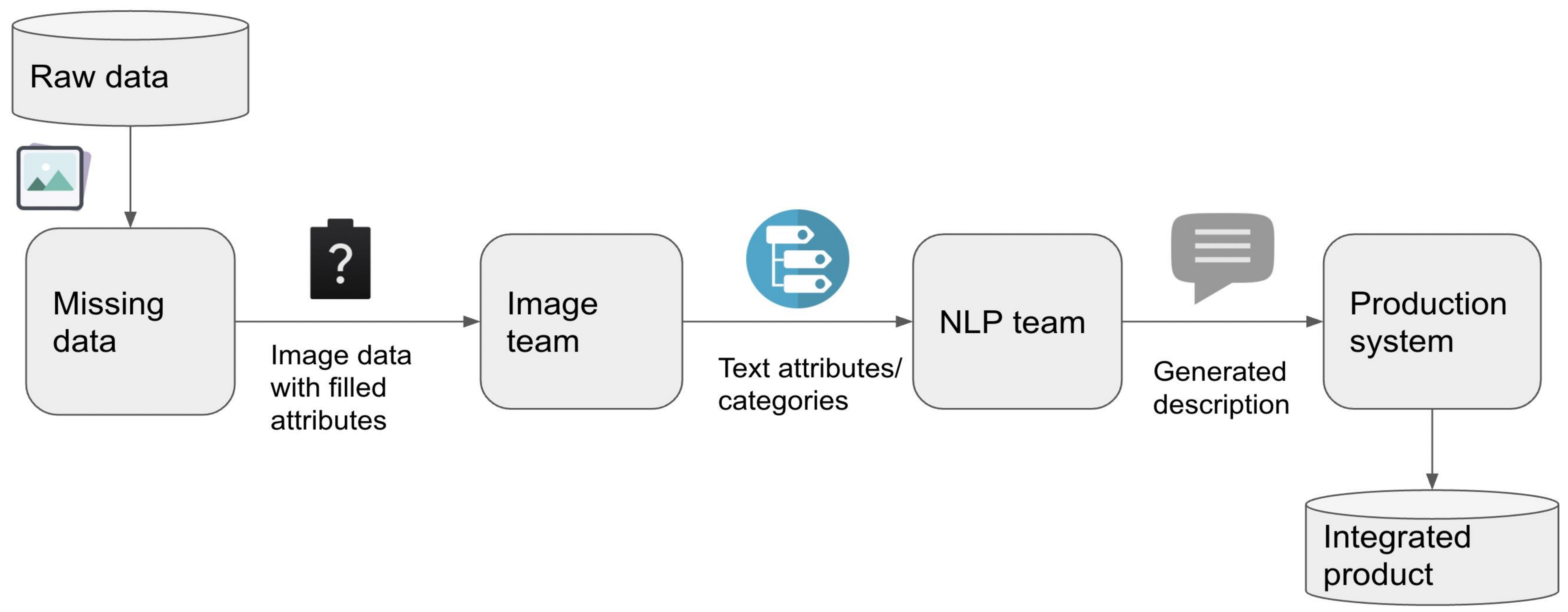
PURDUE UNIVERSITY

General Section

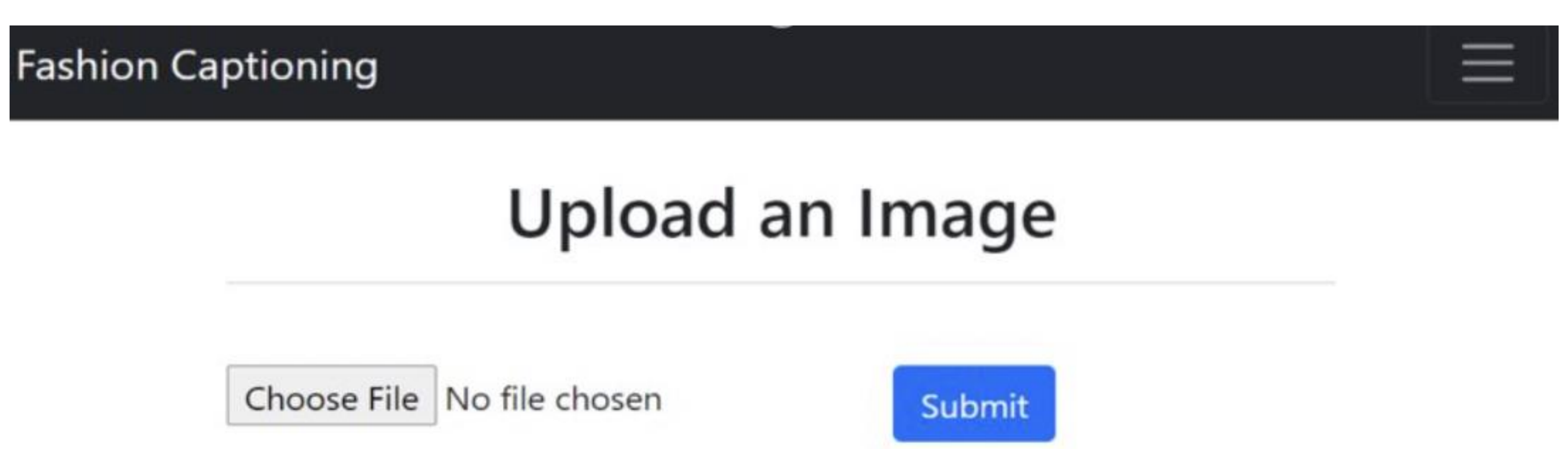
The problem we are working on is replicate and making improvement on the fashion captioning generating pipeline bring out by the Farfetch paper: Fashion Captioning: Towards Generating Accurate Descriptions with Semantic Reward

According to the paper, generating accurate descriptions for online fashion items is important not only for enhancing customers' shopping experiences, but also for the increase of online sales.

The project has break down into 4 sub parts and working by different sub teams. We aim to build an integrated product based on the fashion image generating pipeline as an end result.



Pipeline integrated product demo:



MISSING DATA

The goal of this team is to find missing or poorly annotated data through data profiling and, ultimately, find ways to model the unrepresented information/knowledge for product titles, categories and attributes.

Generating missing attributes given initial distribution of attributes and categories

Demo: **Attributes**: ["diamond", "black", "dial", "watch", "steel", "high", "tech", "ceramic"]

New Attributes: ["diamond", "black", "dial", "watch", "steel", "high", "tech", "ceramic", "shiny", "luxurious"]

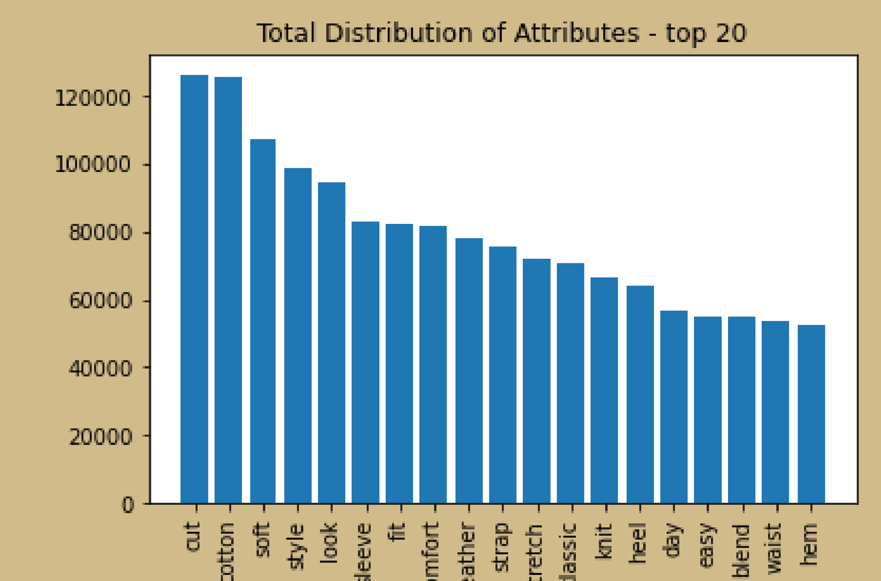
Add missing attributes that accurately describe a product through means of data cleaning, word embeddings, clust ering.

Metadata Extraction: Search through metadata to find any additional attributes that were left out or removed in the initial data reduction. Attributes specifically located in product titles

Word Embedding: Use network with 3 layers to find relationship and output 100-dimensional array using CBOW method.

Sum of Distance: Use word embedding distance to extract missing attributes. Given each attribute in the original list of attributes, extract the top 10 most similar words for each. Concatenate all sets of 10 words and sum their embedding. Return the attribute with the largest sum.

Evaluation: Created our own testing dataset where we pulled out attributes from the dataset to see if our model could predict those attributes that we know are correct.



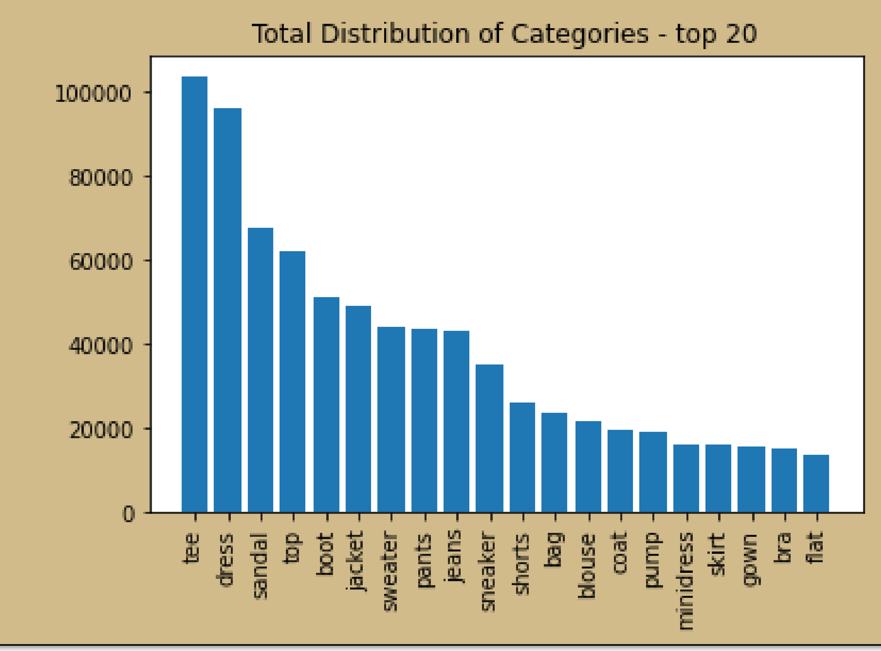


IMAGE-PROCESS TEAM

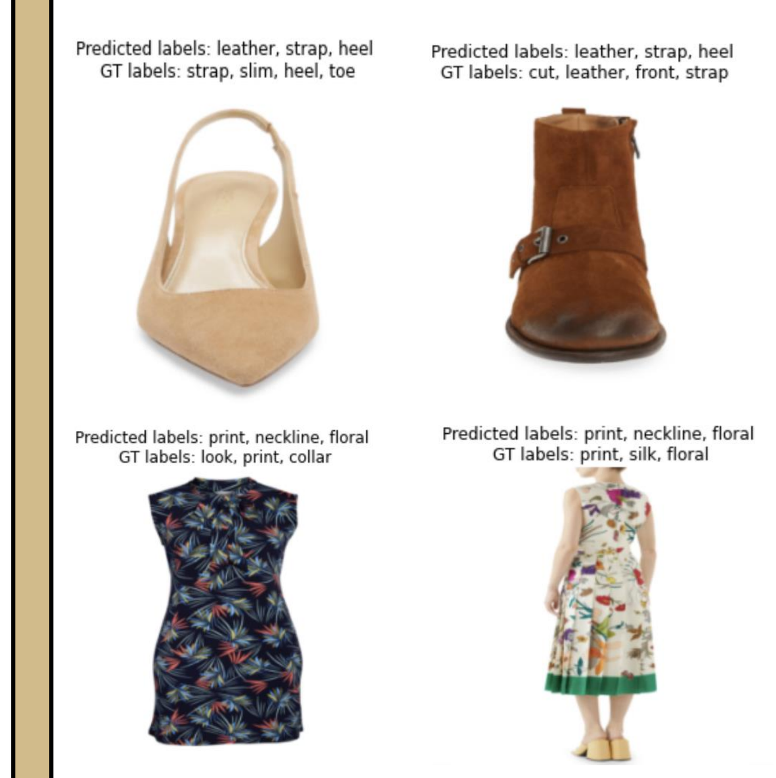
This team is aiming to build classification models for image categories and attributes. The text output will be used by NLP team to generate description sentences.

Multi-label classification
model: generating attributes in text from given images

- Model information:
  - Input: images
  - Output: list of attributes regards to images
  - Loss function: sigmoid + cross entropy loss
- Evaluation:
  - Mean average precision
  - F1 score (micro, macro,sample)
- Training:
  - Select top 50 attributes as Missing data section presented
  - Current evaluation score table:

Epoch	Micro f1	Macro f1	Sample f1
0	0.108	0.065	0.105
1	0.101	0.069	0.098
4	0.019	0.011	0.011
5	0.074	0.037	0.057

Demo images multilabel predictions:



Multi-class classification: classify the fashion images into categories, such as clothing, shoes, accessories, etc.

- Model information:
  - Use model Resnet and CrossEntropyLoss loss function.
  - Apply layer of Softmax at the end to generate probability distribution for each classes
- Evaluation:
  - Precision/Recall

Epoch	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
0	1.998	44.9%	1.4443	57.3%
1	1.461	57.66%	1.2348	63.2%
2	1.2828	62.23%	1.1744	65.2%

After testing the model on 3000 testing samples, we got

Average precision score: 0.61

Average recall score: 0.53

Demo images multiclass predictions:

FASHION IMAGE CAPTIONING (NLP/Production System)

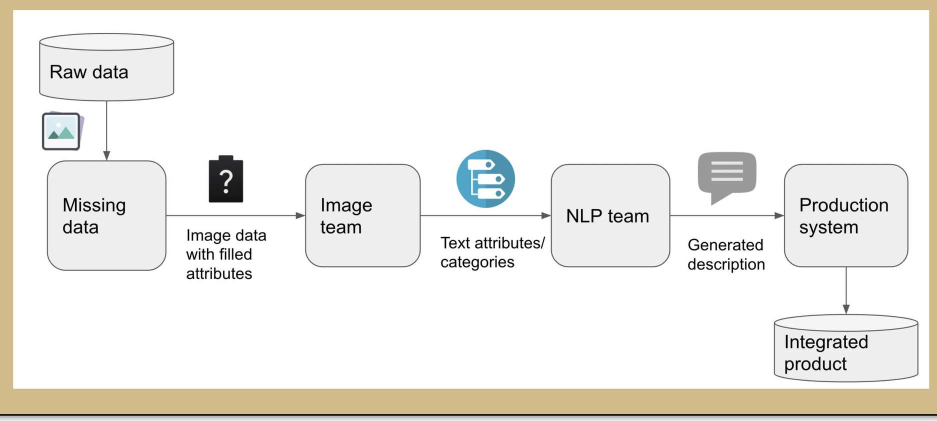
In collaboration with FARFETCH

H. Gu, H. Henderson, B. Kerr, A. Kotalwar, A. Lakkad, J. Moore, S. Nattuvetty, S. Ravi, J. Rusboldt, H. Wan

General Section

The problem we are working on is replicate and making improvement on the fashion captioning generating pipeline bring out by the Farfetch paper: Fashion Captioning: Towards Generating Accurate Descriptions with Semantic Reward

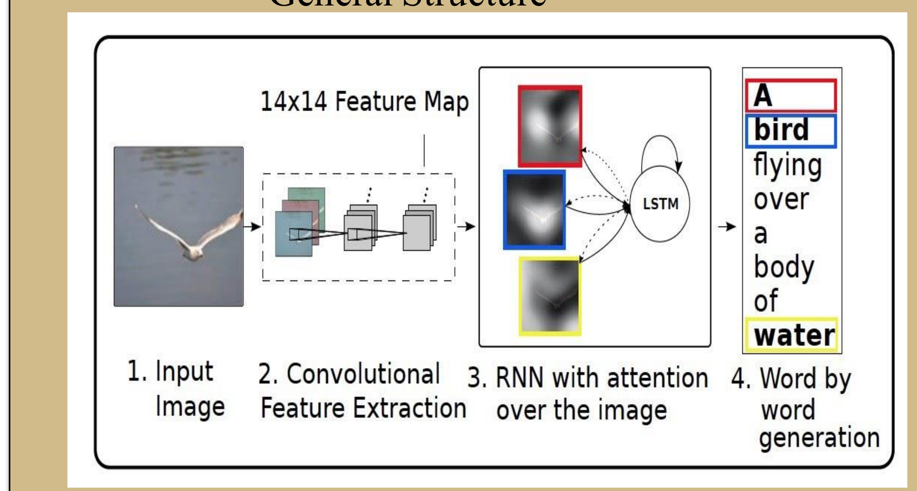
According to the paper, generating accurate descriptions for online fashion items is important not only for enhancing customers' shopping experiences, but also for the increase of online sales.
The project has break down into 4 sub parts and working by different sub teams
We aim to build an integrated product based on the fashion image generating pipeline as an end result.



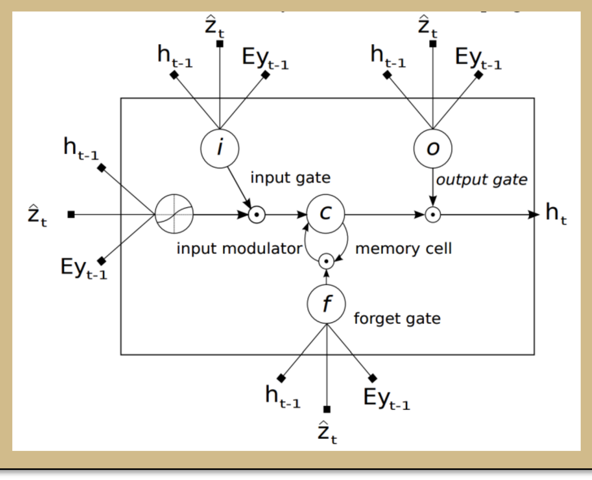
NATURAL LANGUAGE PROCESSING (NLP)

NLP Team Goal:
(1) Generate image captions
(2) Use extracted image embeddings from Image Processing Team, as SAT input (Show, Attention, and Tell)
(3) Improve generated captions: make them similar to actual captions

- **Model Information:**
  - General Structure



  - A LSTM Cell



- **Demonstration**:

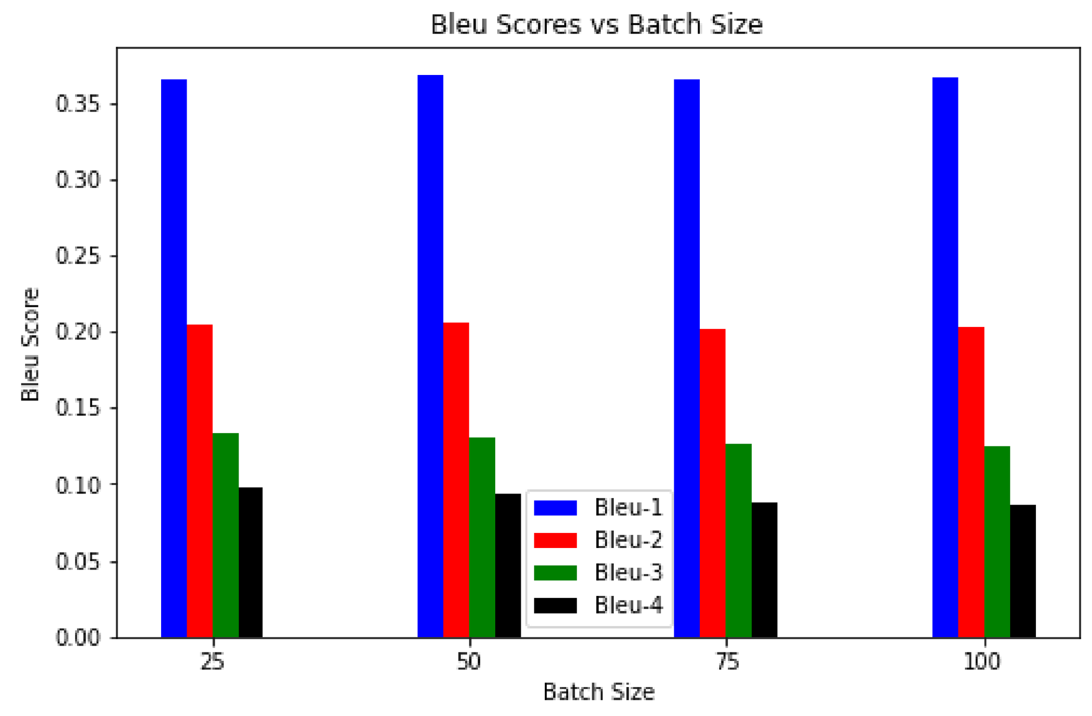
- **Prediction**: a little spin on summer day or evening look in this lace top trimmed with eyelash fringe that delicately frame sun kissed shoulder
- **Reference**: put a romantic spin on summer day or evening look in this lace top trimmed with eyelash fringe that delicately frame sun kissed shoulder

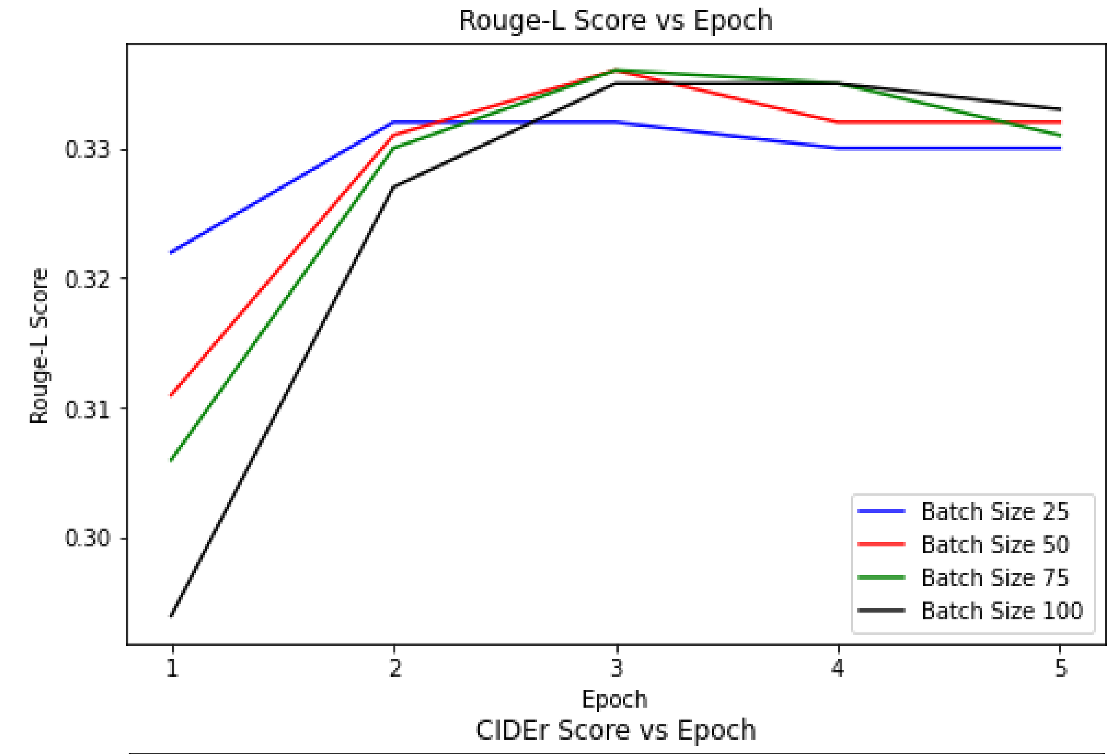
- **Evaluation**:
  - BLEU (Bilingual Evaluation Understudy)
  - Rouge-L
  - CIDEr (Consensus-based Image Description Evaluation

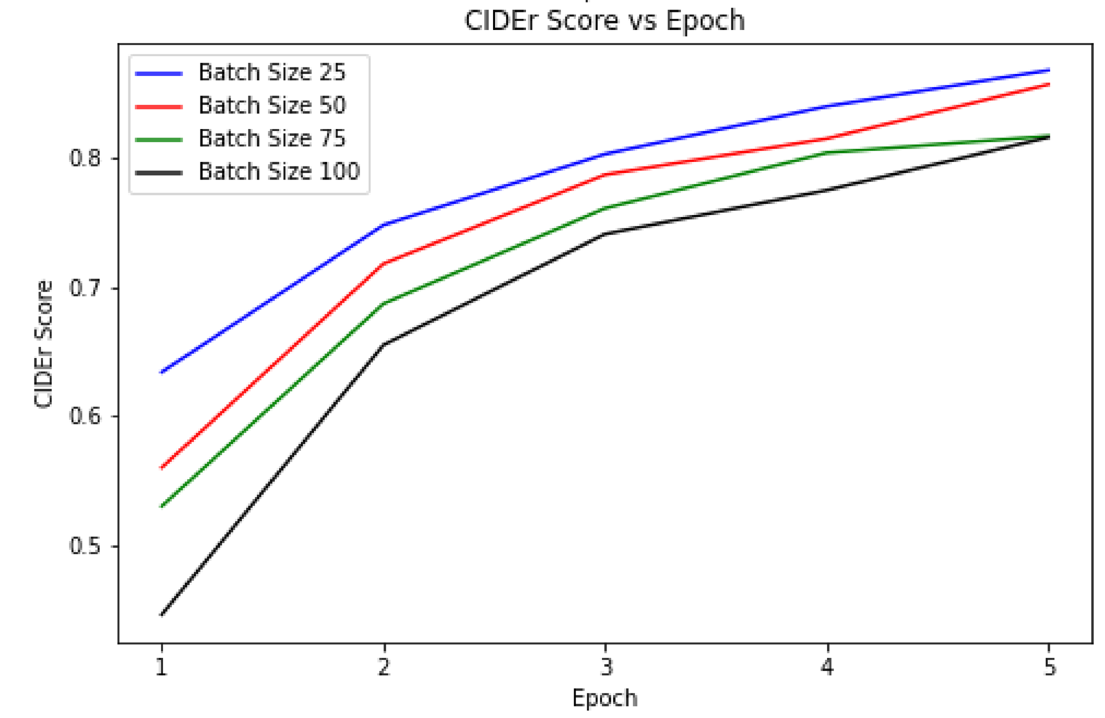
NATURAL LANGUAGE PROCESSING (NLP)

Evaluation Metrics:

- **BLEU** (Bilingual Evaluation Understudy)
  - Average of n-gram precisions between candidates and reference captions
- **Rouge-L**
  - Longest Common Subsequence (LCS)
  - Helps evaluate the existence of repeated attribute details in a generated caption
- **CIDEr** (Consensus-based Image Description Evaluation)
  - TFIDF-based metric
  - Calculates how well candidate sentences matches the consensus of a set of image descriptions





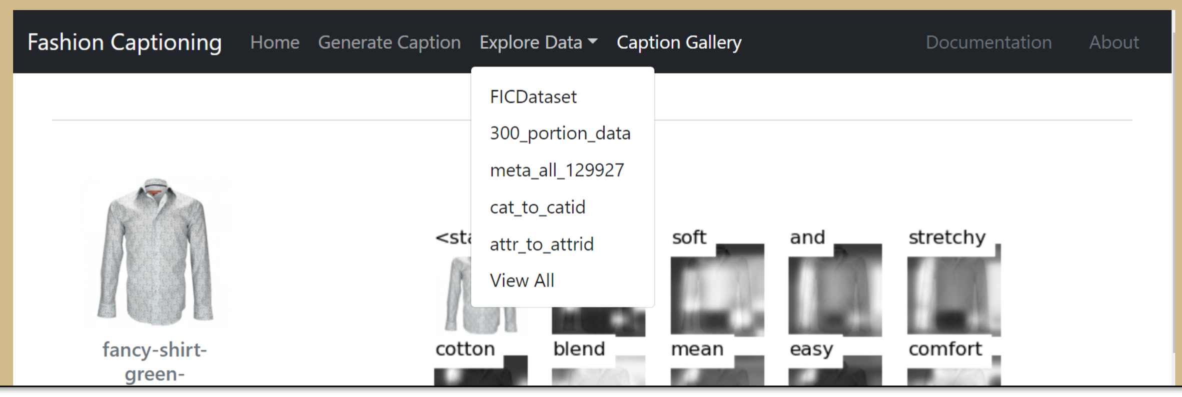


PRODUCTION SYSTEM

Team Goals:
- Develop a full-stack and deployment-ready web application for captioning model
- API and interactive front-end with Python back-end

Implementation:
- Python's Flask library for extensibility and API/site creation
- Celery task queue implemented with Redis server for asynchronous capability
- SQLAlchemy database for scalable image/caption/info storage
- Flask 'templates' filled in via Python data to return front-end HTML pages
- Mobile-friendly, styled in Bootstrap for clean HTML/CSS

Future Improvements:
- RESTful API returning only JSON
- AJAX/jQuery for client browser to update HTML template page with results when asynchronous requests are completed
- Further integration with Celery to avoid freezing/long load times for large requests
- Revamp dataset explorer section of API/website



CONCLUSION

- **Future Goals**
  - Explore the impact of inferencing context while generating caption
  - Train more epochs for image classification models and fine tune hyperparameters
  - Implement a preprocessing procedure to bound the target item on the training image
  - Impact of stemming attributes in image data vs stemming in the NLP step
  - Fine-tuning our system pipeline with images of Farfetch's products

- **References**

  - Yang, X., Zhang, H., Jin, D., Liu, Y., Wu, C., Tan, J., Wang, X. (2020, August 06). Fashion captioning: Towards generating accurate descriptions with semantic rewards. Retrieved February 15, 2021, from https://arxiv.org/abs/2008.02693
  - Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R. & Bengio, Y. (2015). Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. *Proceedings of the 32nd International Conference on Machine Learning*, in *Proceedings of Machine Learning Research* 37:2048-2057 Retrieved February 14, 2021, from http://proceedings.mlr.press/v37/xuc15.html

ACKNOWLEDGEMENTS

We would like to thank our Farfetch Corporate Partner Mentors Ricardo Sousa, Pedro Miguel Ferreira, and Erika Nitsch for providing us this opportunity and guiding us as we created and refined this project. We would also like to thank Dr. Mark Daniel Ward, Ellen Gundlach, Kevin Amstutz, Justin Gould, and Maggie Ann Betz for providing the resources and support for the successful completion of our project.