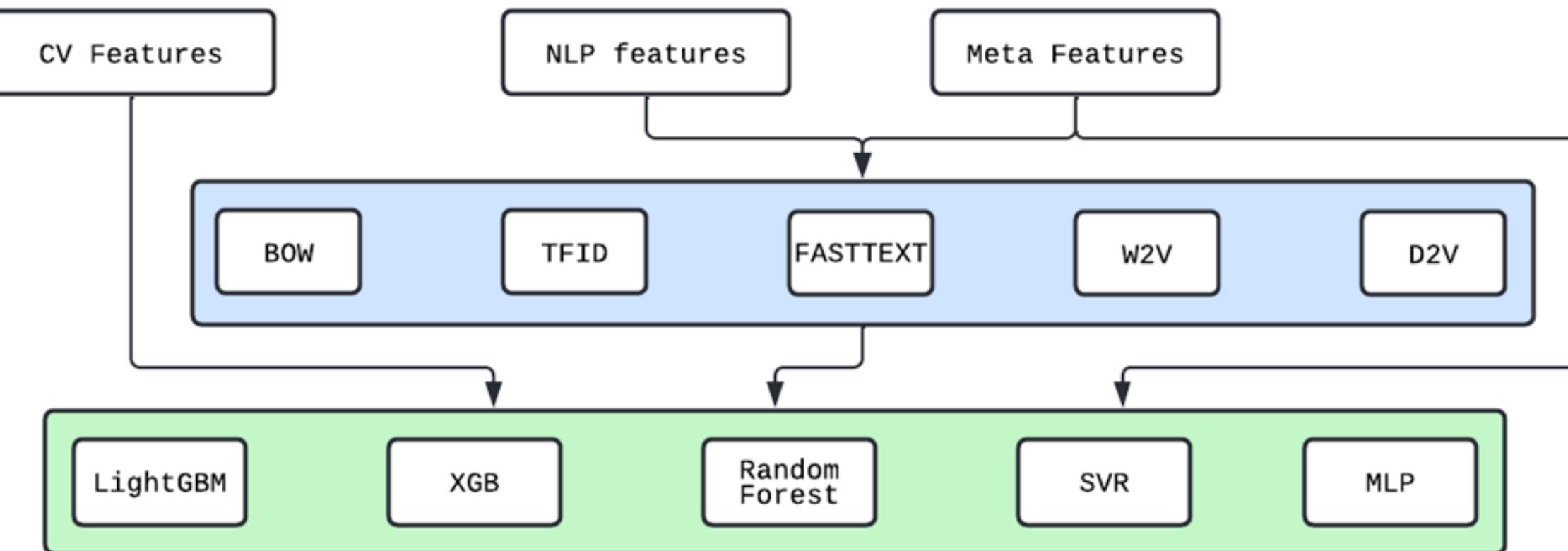
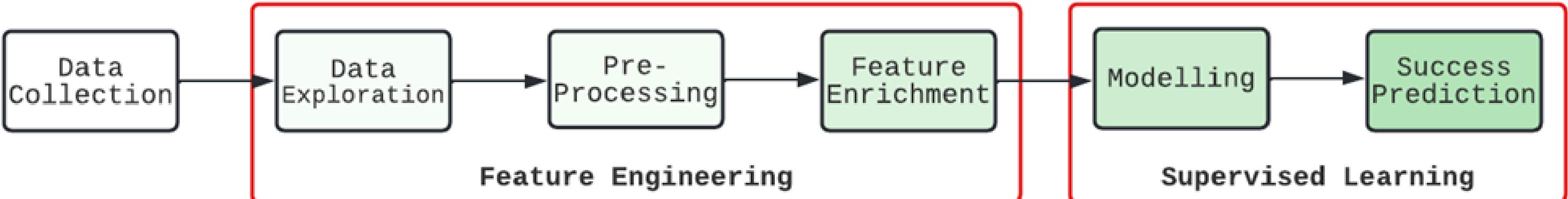


# Success Prediction of Instagram Posts

Sandeep Kumar Kushwaha  
Jyoti Yadav  
Zarmina Ursino

# Workflow



# Data Collection

## META-FEATURES

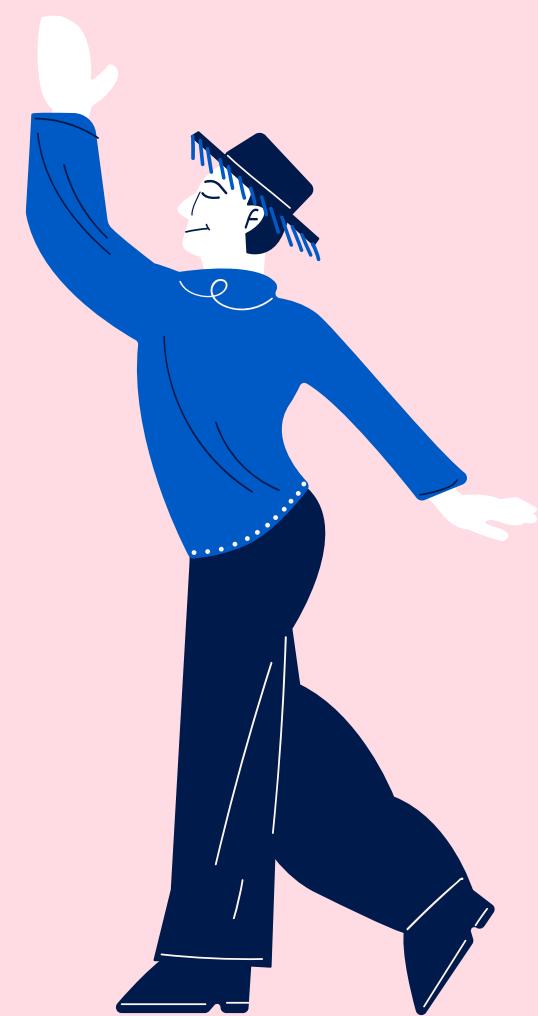
caption\_length  
biography\_length  
po\_co  
pof  
user\_count  
profile\_name\_len  
fol  
act  
pos  
comments\_max  
comments\_min  
comments\_mean  
comments\_std  
comments\_followers  
followers\_comments\_mean  
fol2  
fol\_pos  
fol\_pow  
po\_co\_pow  
comments\_mean\_diff  
Dow  
Hod  
Date

## NLP FEATURES

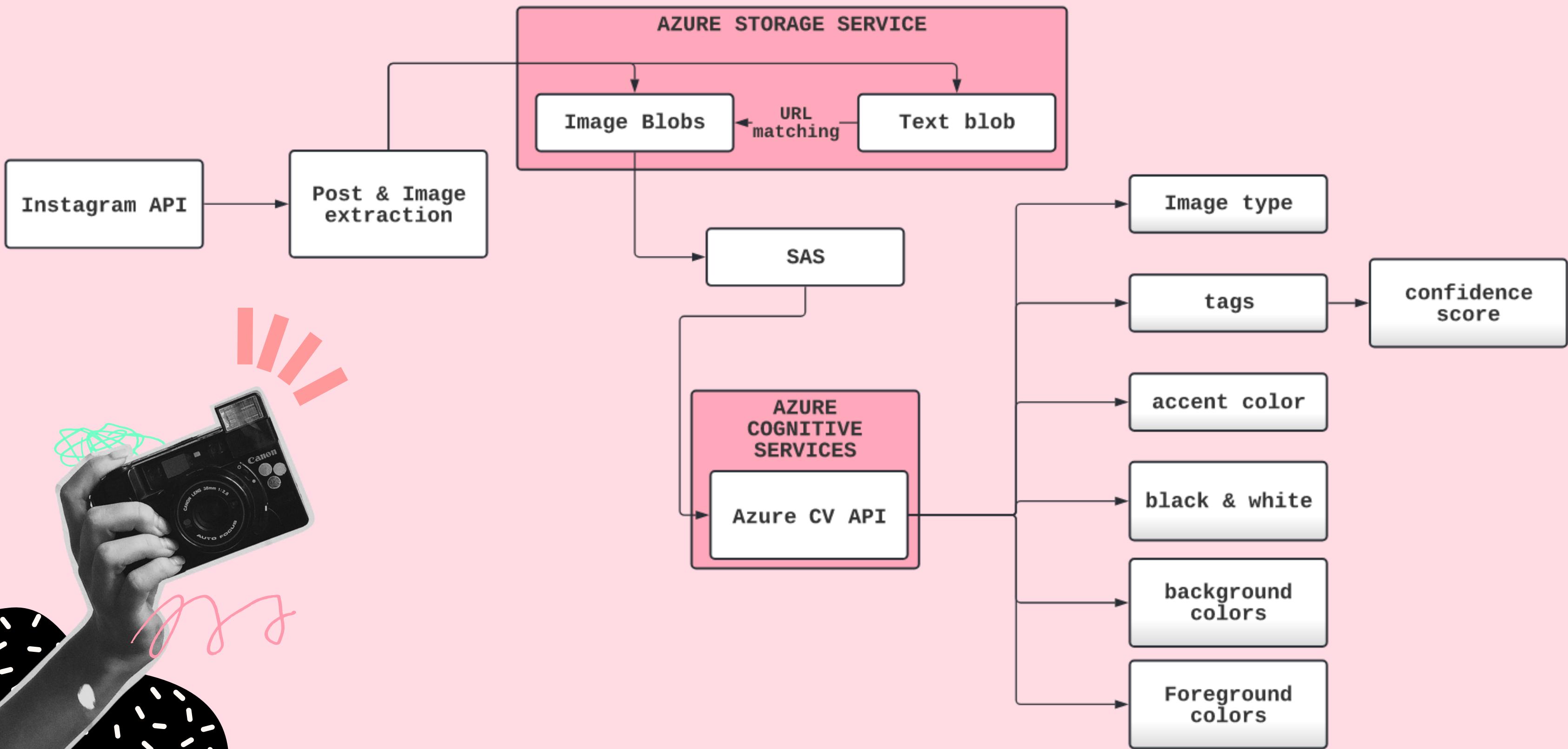
Caption  
caption\_length  
Textblob\_subjectivity  
Textblob\_polarity  
hashtags  
hashcounts  
mentions  
mention\_count  
emoji  
emoji\_count  
emoji\_text  
hashtag\_popularity

## CV FEATURES

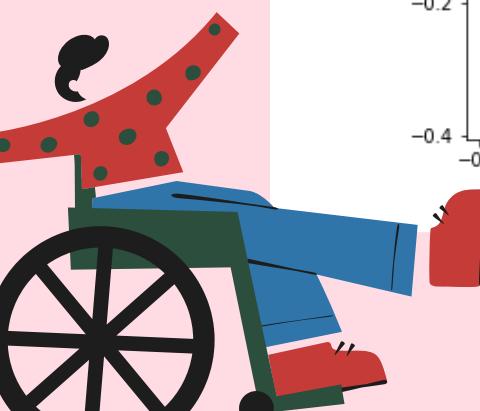
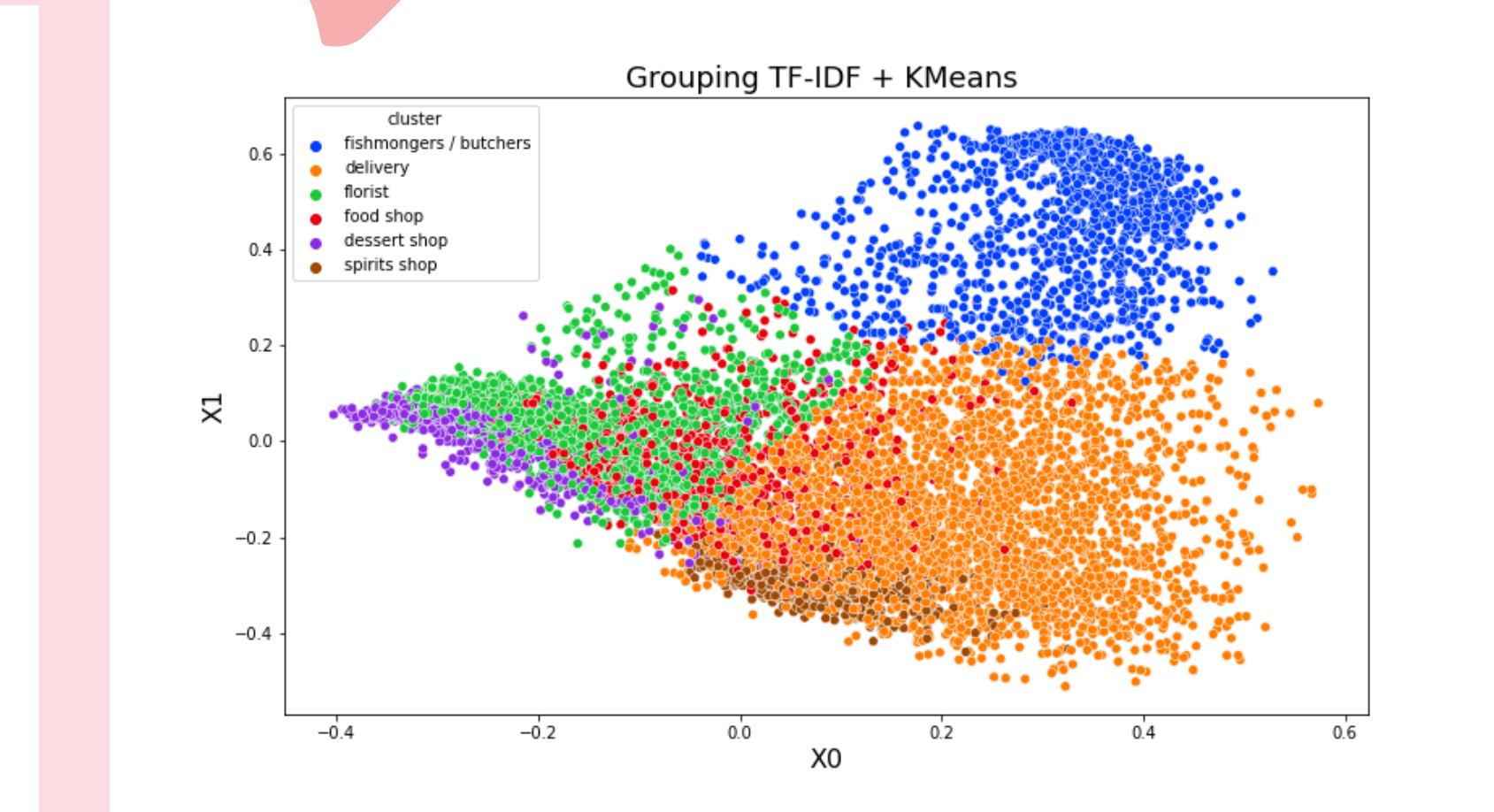
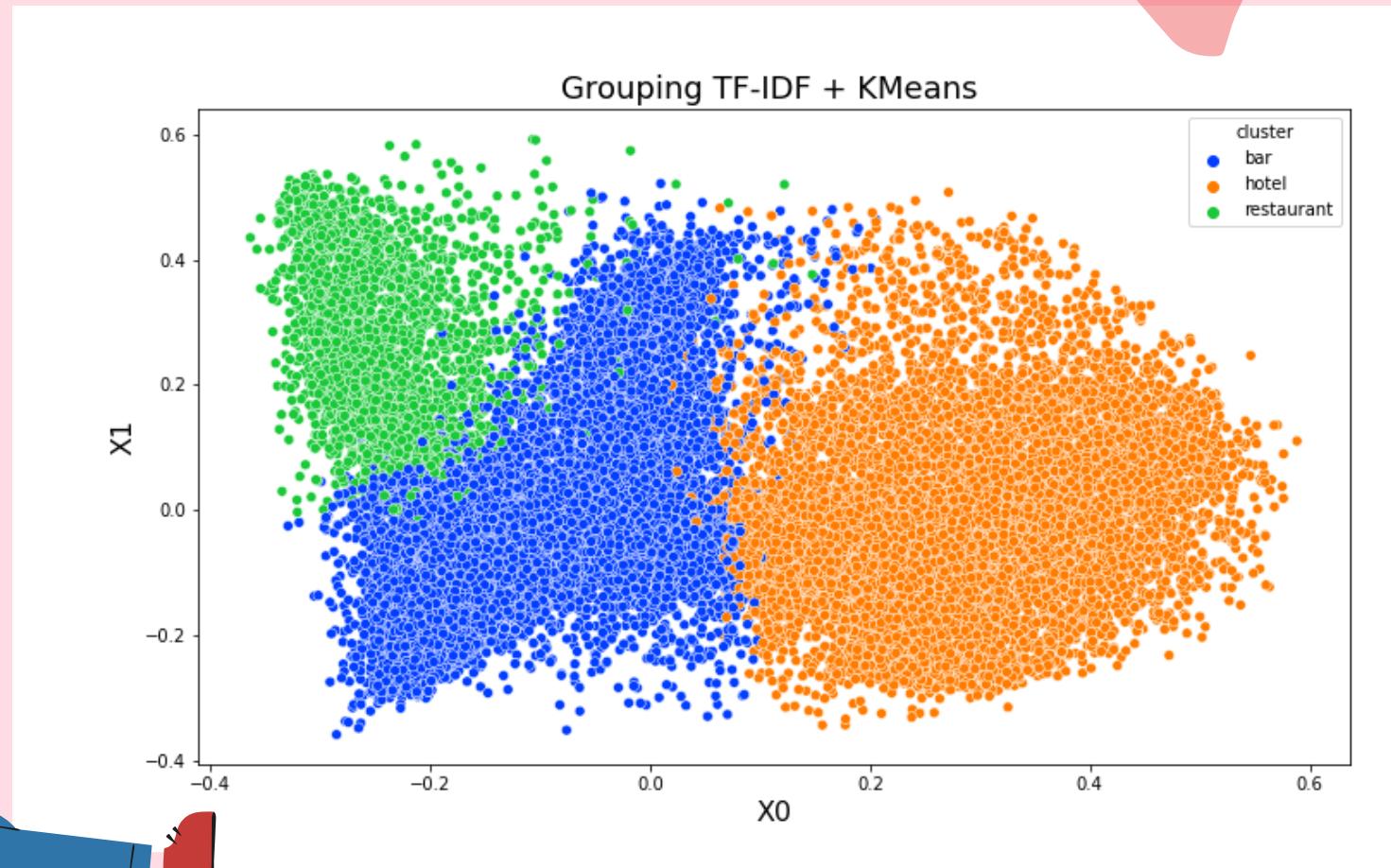
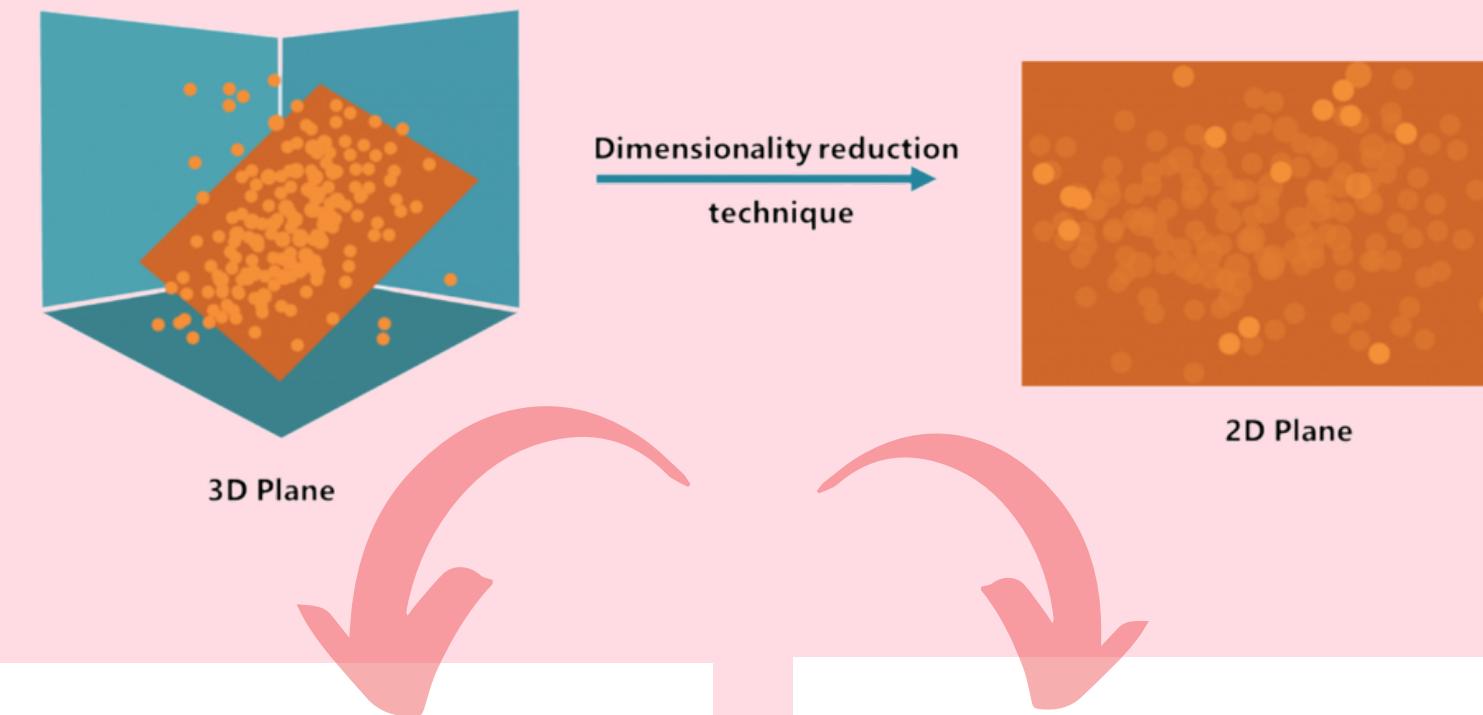
Image type  
Tags  
Confidence score  
Accent colour  
Is\_bw  
Dominant\_colors  
Bg\_colors  
Fore\_colors



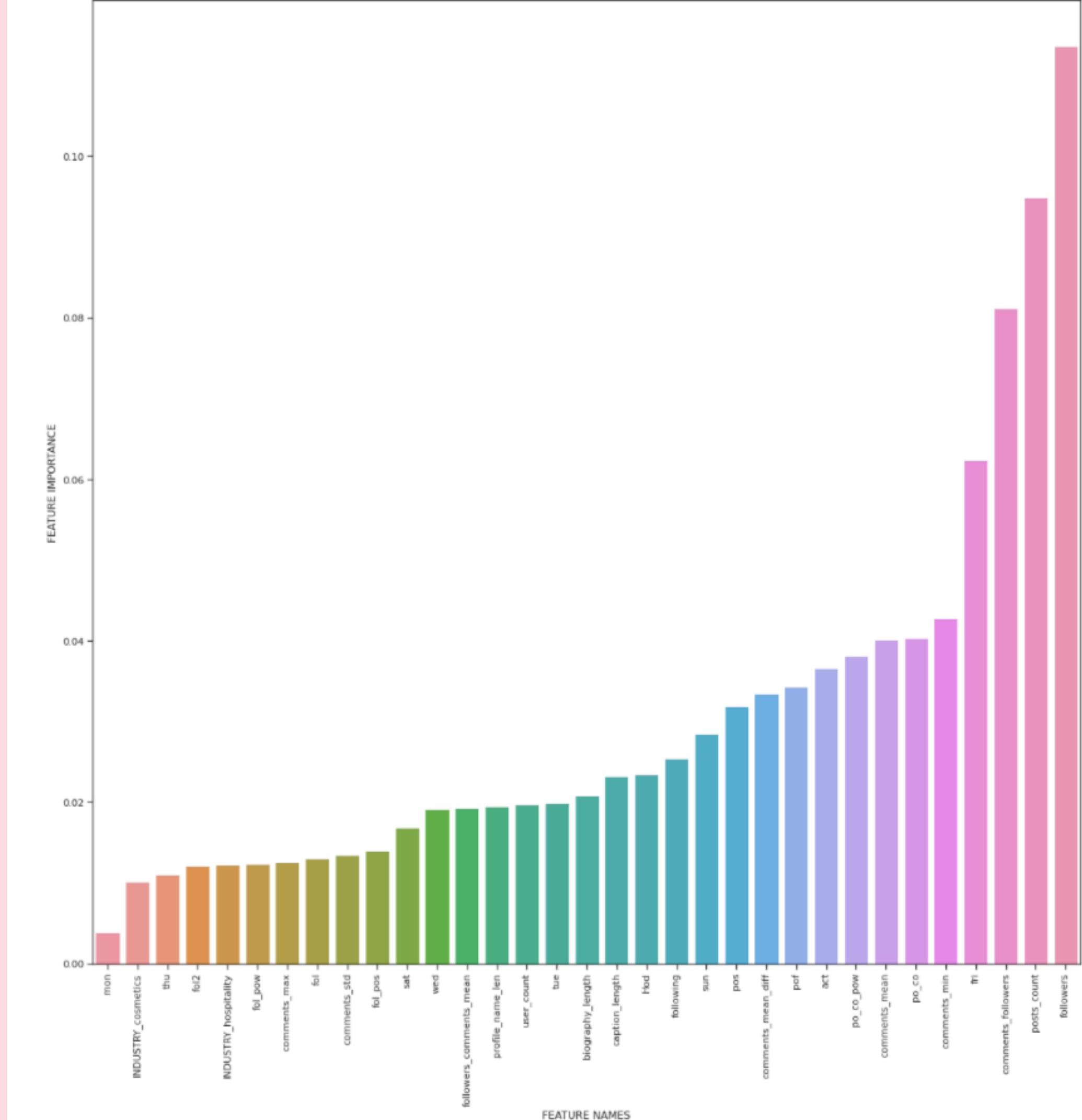
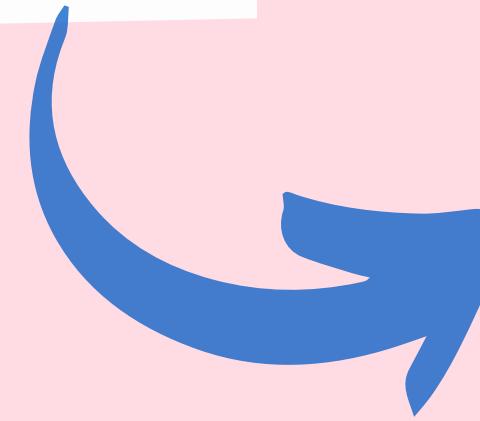
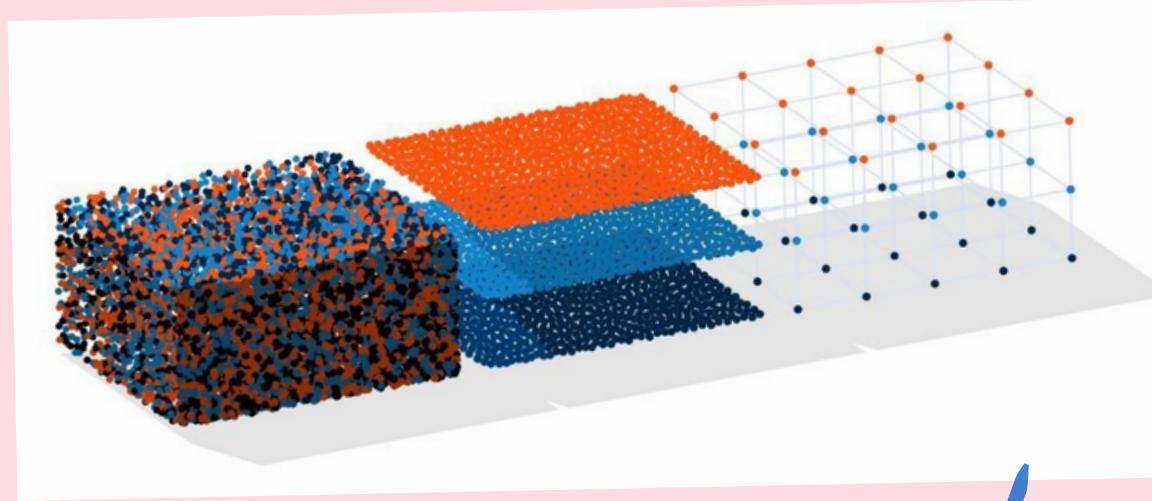
# Extraction of CV features



# Data Exploration - Clustering



# MetaFeatures - Feature Engineering



# Algorithms overview

## LightGBM

- Higher accuracy and a faster training speed.
- Low memory utilization
- Good accuracy and handles overfitting much better while working with smaller datasets.
- Parallel Learning support  
Compatible with both small and large datasets

## XGBoost

- Data scientist actively contributing to XGBoost open-source development
- Cloud integration that supports AWS, Azure, Yarn clusters, and other ecosystems
- Active production use in multiple organizations across various vertical market areas
- A library that was built from the ground up to be efficient, flexible, and portable

## Random Forest

- Well-suited for both regression & classification problems
- It handles missing values and maintains high accuracy
- It can handle very large data sets , making it a good tool for dimensionality reduction.

## SVM

- It can be used for the data with irregular distribution
- By the application of associated kernel function we can solve any complex problem
- Input data can be converted into High dimensional data

## MLP

- Unlike other classification algorithms such as Support Vectors, MLPClassifier relies on an underlying Neural Network to perform the task of classification.



# Parameters

	LGBM	XGB	RF	SVR	MLP
<b>Activation</b>					relu 0,002
<b>alpha</b>					
<b>column sample by tree</b>	0,8	0,7		True	
<b>dual</b>					
<b>Layer size</b>					(50,50,50)
<b>C</b>				1,0	
<b>loss</b>				epsilon_intensive	
<b>epsilon</b>				1,5	
<b>Intercept scaling</b>				1,0	
<b>solver</b>					Adam
<b>Lr</b>	0,01	0,03			constant
<b>Max depth</b>	8	5	10		
<b>Min child weight</b>	1	4			

	LGBM	XGB	RF	SVR	MLP
<b>Min split gain</b>	0,0222				
<b>Min Sample Split</b>			10		
<b>Min sample leaf</b>			4		
<b>N estimators</b>	35000	500	20		
<b>Max features</b>			sqrt		
<b>Max iter</b>					1000
<b>Tol</b>					0,0001
<b>Num leaves</b>	966				
<b>Reg alpha</b>	0,04				
<b>Reg lambda</b>	0,073				
<b>Subsample</b>	0,6	0,7			
<b>Number thread</b>			4		
<b>Verbose</b>					0
<b>Bootstrap</b>				True	

# Word Representations for NLP and CV features

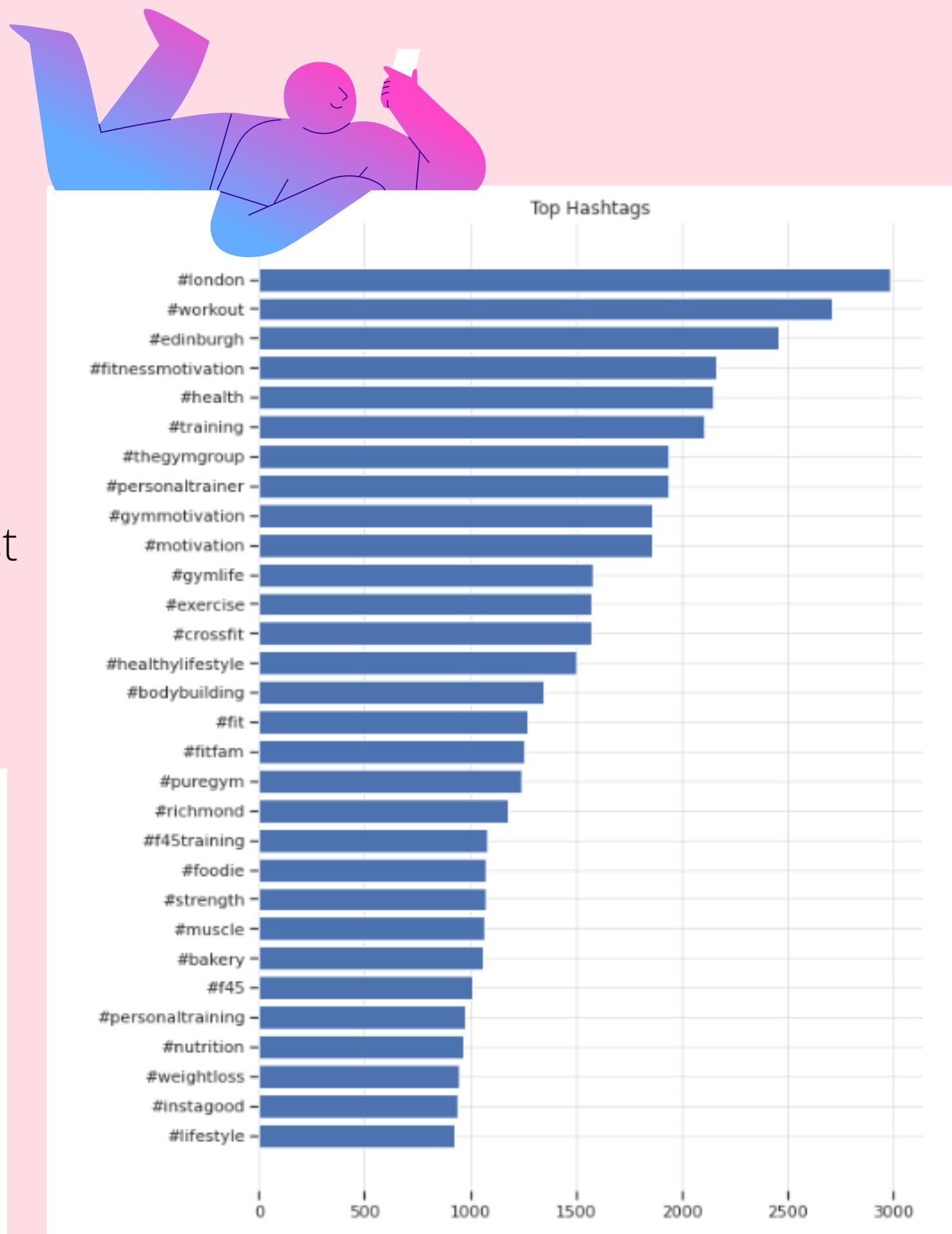
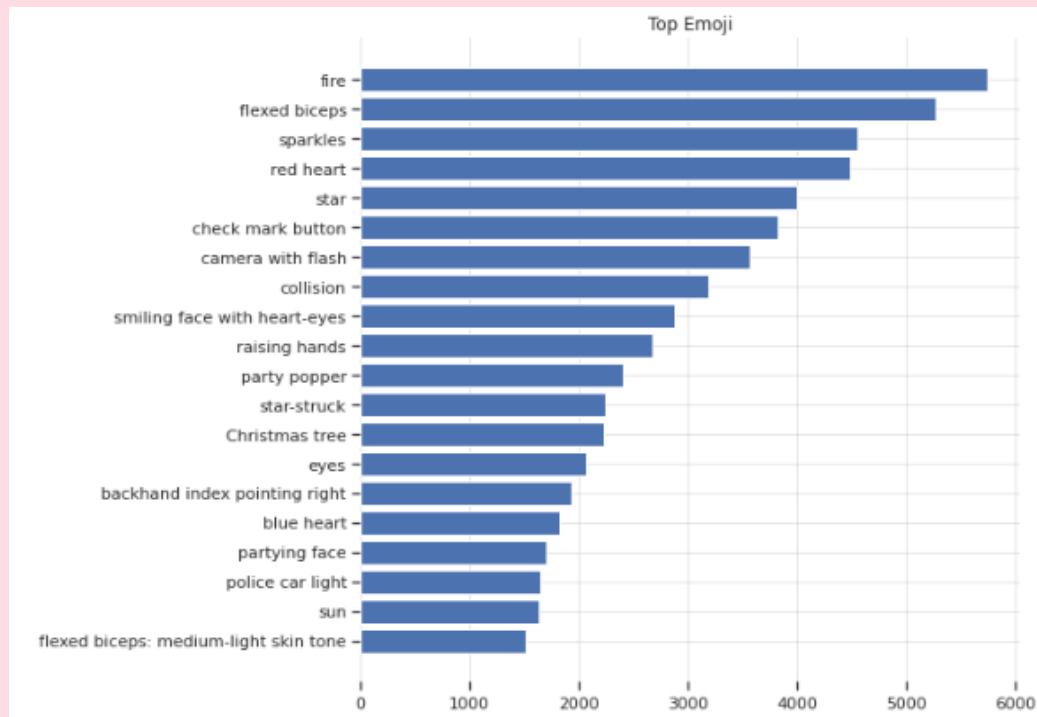
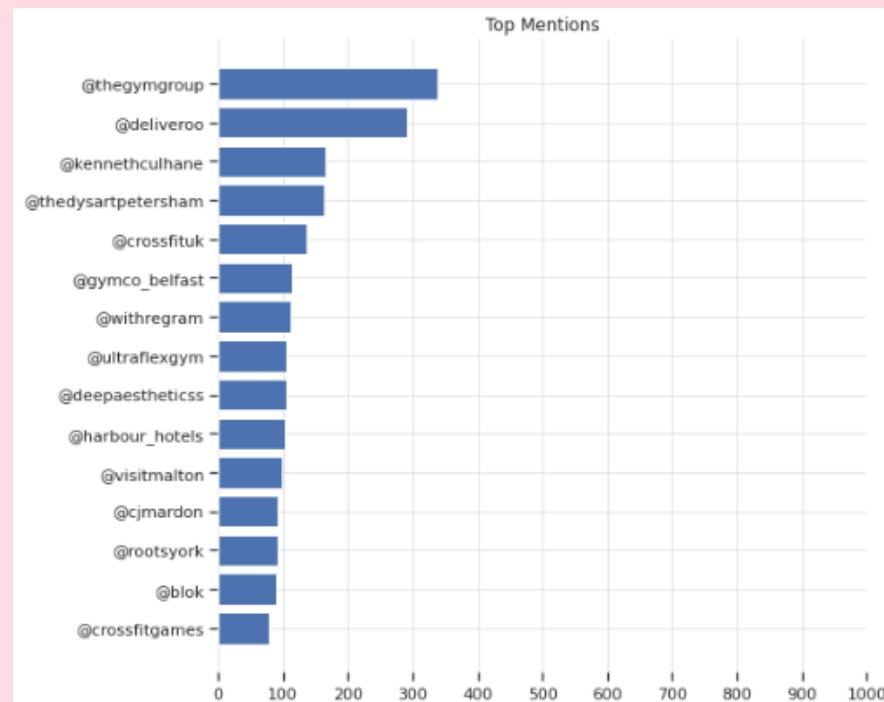
- Bag of Words
- TF-TFIDF
- Word2Vec
  - CBOW
  - SkipGram
- Doc2Vec
- FastText
  - CBOW
  - SkipGram



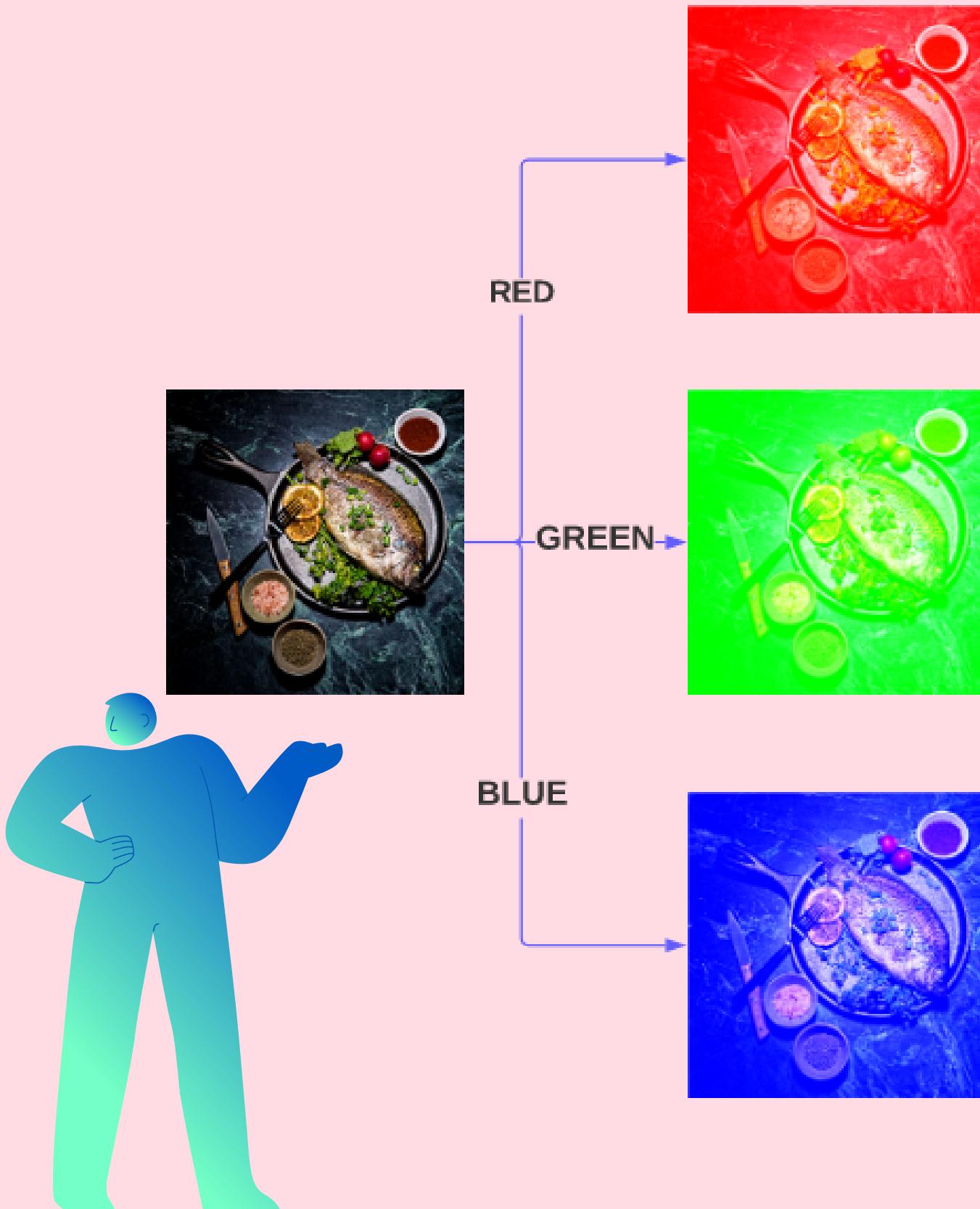
# NLP features - pre processing

Preprocessing on captions

- lowercase
- remove mentions
- remove hashtags
- replace special characters with space
- remove brackets
- removes any special character that is not in good symbols list
- remove stopwords
- removes any left or right spacing



# CV features - preprocessing



We performed the following preprocessing on raw data:

**For images:** we converted the string values to list of RGB values. Each color is converted into three columns - R, G, B values.

**For textual data:** we performed regex operations to remove unnecessary parts and convert the tags with high confidence into a separate list that is later handled by FastText.

# Analysis and Conclusions



Metrics	Base Model	BOW	TF-IDF	Word2Vec Skip gram	Word2Vec CBOW	Fast Text Skip Gram	Fast Text CBOW	Doc2vec	CV
<b>LightGBM</b>									
<b>RMSE</b>	0.008620	0.008644	0.008625	0.007498	0.007447	0.007131	0.007084	0.009629	0.008122
<b>R2 Score</b>	0.577815	0.582151	0.583974	0.610671	0.621513	0.605669	0.610846	0.522325	0.577527
<b>MAE</b>	0.001825	0.001836	0.001838	0.001773	0.001787	0.001858	0.001857	0.001820	0.001765
<b>RMSLE</b>	-4.75371	-4.75086	-4.75304	-4.893115	-4.899883	-4.94323	-4.94984	-4.64295	4.813200
<b>MaxError</b>	0.533904	0.537682	0.538299	0.323809	0.304083	0.197807	0.194891	0.864437	0.494410
<b>XGBoost</b>									
<b>RMSE</b>	0.007876	0.007805	0.008174	0.006736	0.007082	0.006502	0.006267	0.009472	0.007456
<b>R2 Score</b>	0.647521	0.659310	0.626353	0.685829	0.657789	0.672172	0.695508	0.537838	0.643937
<b>MAE</b>	0.001464	0.001491	0.001510	0.001435	0.001437	0.001472	0.001466	0.001472	0.001377
<b>RMSLE</b>	-4.84393	-4.85293	-4.80676	-5.000359	-4.950260	-5.03558	-5.07251	-4.65945	-4.89871
<b>MaxError</b>	0.520647	0.484182	0.526286	0.234300	0.293903	0.183032	0.174488	0.889369	0.479147



### Random Forest

<b>RMSE</b>	0.007606	0.008736	0.008484	0.007624	0.007576	0.006998	0.007174	0.009653	0.009934
<b>R2 Score</b>	0.671264	0.573200	0.597485	0.597488	0.608360	0.620265	0.600948	0.520011	0.368012
<b>MAE</b>	0.001413	0.001714	0.001640	0.001672	0.001704	0.001717	0.001761	0.001690	0.002498
<b>RMSLE</b>	-4.87880	-4.74026	-4.76955	-4.876465	-4.882802	-4.96209	-4.93728	-4.64053	-4.61182
<b>MaxError</b>	0.509198	0.578521	0.533026	0.389754	0.271130	0.204303	0.244427	0.866844	0.605138

### SVR

<b>RMSE</b>	0.013675	0.013762	0.013762	0.012425	0.012539	0.011823	0.011823	0.011841	0.012907
<b>R2 Score</b>	-0.06264	-0.05914	-0.05914	-0.069177	-0.072932	-0.08390	-0.08390	0.277668	-0.06690
<b>MAE</b>	0.003320	0.003252	0.003252	0.003161	0.003269	0.003290	0.003290	0.003542	0.003232
<b>RMSLE</b>	-4.29217	-4.28581	-4.28581	-4.388005	-4.378898	-4.43767	-4.43767	-4.43617	-4.35000
<b>MaxError</b>	0.639221	0.691846	0.691846	0.523607	0.387544	0.363963	0.363963	0.960571	0.639221

### MLP

	0.009841	0.010085	0.010117	0.009621	0.009720	0.008891	0.009380	0.011841	0.010449
<b>RMSE</b>	0.009841	0.010085	0.010117	0.009621	0.009720	0.008891	0.009380	0.011841	0.010449
<b>R2 Score</b>	0.449671	0.431302	0.427629	0.359014	0.355222	0.387016	0.317820	0.277668	0.300733
<b>MAE</b>	0.002749	0.002363	0.002454	0.003246	0.003136	0.003318	0.003789	0.003542	0.002813
<b>RMSLE</b>	-4.62117	-4.59674	-4.59352	-4.643823	-4.633520	-4.72266	-4.66918	-4.43617	-4.56124
<b>MaxError</b>	0.561765	0.639808	0.641612	0.484641	0.354472	0.290263	0.305460	0.960571	0.591780



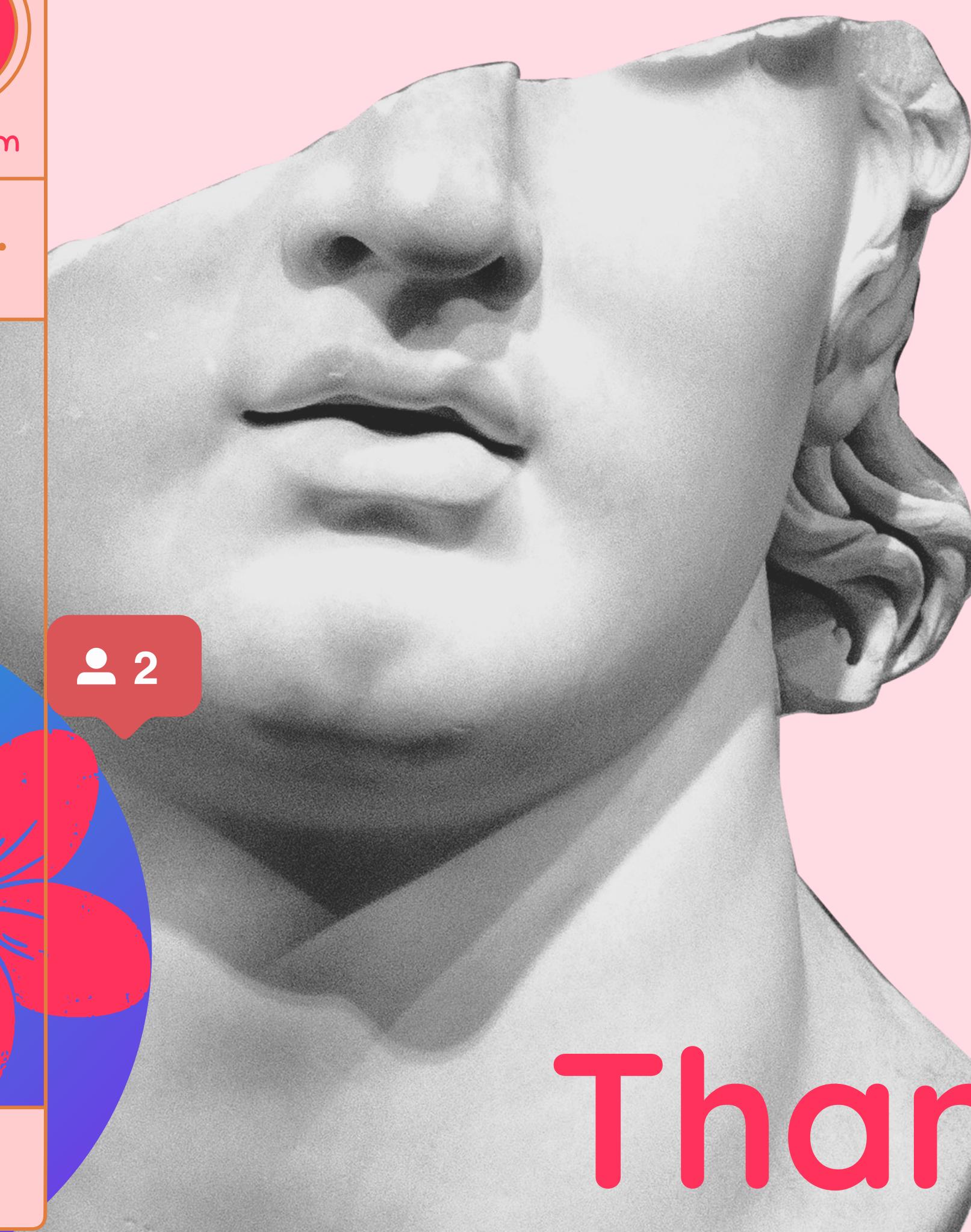


hansolo ella.vator maesharona catmlem



dww\_sum

...



Thank you