

Reviews' Sentiment

Group Project

Prepared By: Biken K.C. & Duy Khanh Pham

Background

On the internet, there are many products listed for sale. For people it is a hassle to manually check the entire reviews of the product manually before purchasing any product. It gonna be lot easier if there is a program that can generate customer's sentiment of any product at once from any platform. This is possible by scraping the reviews of a certain product from a certain platform and passing it through the sentimental analysis model to return a ratio of good reviews as opposed to bad reviews.

Sentiment Analysis

Task: Predict the sentiment of various products from various websites/platform with the help of reviews

Gaming products: Steam

Restaurants: Yelp

Future Topics:

Shopping products: Amazon

Books: Amazon

Movies: IMDB etc.

Dataset

- Gaming and Restaurant from Kaggle
- Gaming is 2 gigabytes
 - Around 9.5 Million data points
 - 2 features i.e. review as text and target label
 - Source: kaggle
- Restaurant is 400 megabytes
 - Around 560 thousands data point
 - 2 features i.e. review as text and target label
 - Source: kaggle

Technique

- Natural Language Processing
- Use of two deep learning model
 - GloVe Model (Global Vectors for Word Representation)
 - BERT Model (Bidirectional Encoder Representations from Transformers)
- Working mechanism:
 - 2 inputs:
 - A certain restaurant or game name
 - A certain platform (for example: steam)
- Based on the input, scraper code will scrape reviews into a csv file
- File will be ran through the model for prediction
- Returns a ratio of good:bad reviews

	app_id	app_name	review_id	language	review	timestamp_created	timestamp_updated	recommended	votes_helpful	votes_funny	...
0	292030	The Witcher 3: Wild Hunt	85185598	schinese	不玩此生遗憾，RPG游戏里的天花板，太吸引人了	1611381629	1611381629	True	0	0	...
1	292030	The Witcher 3: Wild Hunt	85185250	schinese	拔DIAO无情打桩机--杰洛特!!!	1611381030	1611381030	True	0	0	...
2	292030	The Witcher 3: Wild Hunt	85185111	schinese	巫师3NB	1611380800	1611380800	True	0	0	...
3	292030	The Witcher 3: Wild Hunt	85184605	english	One of the best RPG's of all time, worthy of a...	1611379970	1611379970	True	0	0	...
4	292030	The Witcher 3: Wild Hunt	85184287	schinese	大作	1611379427	1611379427	True	0	0	...
...
21747366	546560	Half-Life: Alyx	65645125	english	Holy balls	1584986704	1584986704	True	0	0	...
21747367	546560	Half-Life: Alyx	65645115	english	A tripod thing bears down on you, it looks imm...	1584986693	1584986693	True	0	0	...
21747368	546560	Half-Life: Alyx	65645100	english	Honestly this is the best vr game ever	1584986673	1584986673	True	0	0	...
21747369	546560	Half-Life: Alyx	65645066	english	Smooth turning is not working right now.vnit a...	1584986631	1586382422	True	0	0	...
21747370	546560	Half-Life: Alyx	65644930	schinese	VWVR加载中闪退，无法进入游戏inVWVR Flashback in Load, cant pl...	1584986505	1592060800	True	0	0	...

Gaming Dataset

```

1 #Unique Languages
2 lang = [x for x in train['language'].unique()]
3 print(lang)
4 print('Total unique languages: ', str(len(lang)))

```

```

['schinese', 'english', 'turkish', 'spanish', 'russian', 'koreana', 'latam', 'brazilian', 'portuguese', 'vietnamese', 'polish',
'french', 'german', 'hungarian', 'ukrainian', 'tchinese', 'bulgarian', 'czech', 'italian', 'thai', 'greek', 'dutch', 'finnish',
'romanian', 'japanese', 'swedish', 'danish', 'norwegian']
Total unique languages: 28

```

```

1 #Unique Features
2 feature = [col for col in train]
3 print(feature)
4 print('Total unique features: ', str(len(feature)))

```

```

['app_id', 'app_name', 'review_id', 'language', 'review', 'timestamp_created', 'timestamp_updated', 'recommended', 'votes_helpful',
'votes_funny', 'weighted_vote_score', 'comment_count', 'steam_purchase', 'received_for_free', 'written_during_early_access',
'author.steamid', 'author.num_games_owned', 'author.num_reviews', 'author.playtime_forever', 'author.playtime_last_two_weeks',
'author.playtime_at_review', 'author.last_played']
Total unique features: 22

```

- Based on the current time crunch we are in, English will only be used
- Only the feature “review” and “recommended” will be used for the training, the rest of the column can be dropped

1	en_train
---	----------

	review	recommended
--	--------	-------------

0	One of the best RPG's of all time, worthy of a...	True
---	---	------

1	good story, good graphics. lots to do.	True
---	--	------

2	dis gud,	True
---	----------	------

3	favorite game of all time cant wait for the Ne...	True
---	---	------

4	Why wouldn't you get this	True
---	---------------------------	------

...
-----	-----	-----

9635432	best game ever\n	True
---------	------------------	------

9635433	Holy balls	True
---------	------------	------


9635434	A tripod thing bears down on you, it looks imm...	True
---------	---	------

9635435	Honestly this is the best vr game ever	True
---------	--	------

9635436	Smooth turning is not working right now.\nlt a...	True
---------	---	------

English Only, Cleared
Gaming Dataset


Massive size reduction for future uses
Reduces time for code to load dataset

 EnOnly_simplified_steam_reviews.csv

4/30/2022 1:02 AM

CSV File

2,129,155 KB

 steam_reviews.csv

1/25/2021 3:29 AM

CSV File



7,980,261 KB

	review	target
0	One of the best RPG's of all time, worthy of a...	1
1	good story, good graphics. lots to do.	1
2	dis gud,	1
3	favorite game of all time cant wait for the Ne...	1
4	Why wouldn't you get this	1
...
9635432	best game ever\n	1
9635433	Holy balls	1
9635434	A tripod thing bears down on you, it looks imm...	1
9635435	Honestly this is the best vr game ever	1
9635436	Smooth turning is not working right now.\nlt a...	1

Boolean Name & Value Changed
Dropped row with NaN value

For consistency with other datasets

Further Reduced

 EnOnly_simplified_steam_reviews.csv	4/30/2022 10:05 PM	CSV File	2,099,834 KB
 steam_reviews.csv	1/25/2021 3:29 AM	CSV File	7,980,261 KB

```
1 rest_train = pd.read_csv('./Datasets/yelp_review_polarity_csv/train.csv')
2 rest_train
```

1 Unfortunately, the frustration of being Dr. Goldberg's patient is a repeat of the experience I've had with so many other doctors in NYC -- good doctor, terrible staff. It seems that his staff simply never answers the phone. It usually takes 2 hours of repeated calling to get an answer. Who has time for that or wants to deal with it? I have run into this problem with many other doctors and I just don't get it. You have office workers, you have patients with medical needs, why isn't anyone answering the phone? It's incomprehensible and not work the aggravation. It's with regret that I feel that I have to give Dr. Goldberg 2 stars.

0	2	Been going to Dr. Goldberg for over 10 years. ...
1	1	I don't know what Dr. Goldberg was like before...
2	1	I'm writing this review to give you a heads up...
3	2	All the food is great here. But the best thing...
4	1	Wing sauce is like water. Pretty much a lot of...
...
559994	2	Ryan was as good as everyone on yelp has claim...
559995	2	Professional \nFriendly\nOn time AND affordabl...
559996	1	Phone calls always go to voicemail and message...
559997	1	Looks like all of the good reviews have gone t...
559998	2	Ryan Rocks! I called him this morning for some...

Restaurant Dataset

review target

0 Unfortunately, the frustration of being Dr. Go... 1

1 Been going to Dr. Goldberg for over 10 years. ... 2

2 I don't know what Dr. Goldberg was like before... 1

3 I'm writing this review to give you a heads up... 1

4 All the food is great here. But the best thing... 2

... ...

559995 Ryan was as good as everyone on yelp has claim... 2

559996 Professional \nFriendly\nOn time AND affordabl... 2

559997 Phone calls always go to voicemail and message... 1

559998 Looks like all of the good reviews have gone t... 1

559999 Ryan Rocks! I called him this morning for some... 2

Fixed

review target

0 Unfortunately, the frustration of being Dr. Go... 0

1 Been going to Dr. Goldberg for over 10 years. ... 1

2 I don't know what Dr. Goldberg was like before... 0

3 I'm writing this review to give you a heads up... 0

4 All the food is great here. But the best thing... 1

... ...

559995 Ryan was as good as everyone on yelp has claim... 1

559996 Professional \nFriendly\nOn time AND affordabl... 1

559997 Phone calls always go to voicemail and message... 0

559998 Looks like all of the good reviews have gone t... 0

559999 Ryan Rocks! I called him this morning for some... 1

Changed target value for consistency
with code

Contrary to other reviews, I have zero complaints about the service or the prices. I have been getting tire service here for the past 5 years now, and compared to my experience with places like Pep Boys, these guys are experienced and know what they're doing. \nAlso, this is one place that I do not feel like I am being taken advantage of, just because of my gender. Other auto mechanics have been notorious for capitalizing on my ignorance of cars, and have sucked my bank account dry. But here, my service and road coverage has all been well explained - and let up to me to decide. \nAnd they just renovated the waiting room. It looks a lot better than it did in previous years.

0	1	Last summer I had an appointment to get new ti...
1	2	Friendly staff, same starbucks fair you get an...
2	1	The food is good. Unfortunately the service is...
3	2	Even when we didn't have a car Filene's Baseme...
4	2	Picture Billy Joel's "Piano Man" DOUBLED mix...
...
37994	1	If I could give 0...I would. Don't do it.
37995	2	Items Selected:\nChocolate Cinnamon Horn\nSmal...
37996	1	Expensive lunch meals. Fried pickles were goo...
37997	1	Highly overpriced and food was cold. Our waitr...
37998	1	I have been using this company for 11 months. ...

Test Dataset Restaurant

	review	target
0	Contrary to other reviews, I have zero complai...	1
1	Last summer I had an appointment to get new ti...	0
2	Friendly staff, same starbucks fair you get an...	1
3	The food is good. Unfortunately the service is...	0
4	Even when we didn't have a car Filene's Baseme...	1
...
37995	If I could give 0...I would. Don't do it.	0
37996	Items Selected:\nChocolate Cinnamon Horn\nSmal...	1
37997	Expensive lunch meals. Fried pickles were goo...	0
37998	Highly overpriced and food was cold. Our waitr...	0
37999	I have been using this company for 11 months. ...	0

Fixed

Games

```
1 print('Recommended reviews: ', len(en_train[en_train['target']==1]))  
2 print('Not recommended reviews: ', len(en_train[en_train['target']==0]))
```

Recommended reviews: 8568127

Not recommended reviews: 1050560

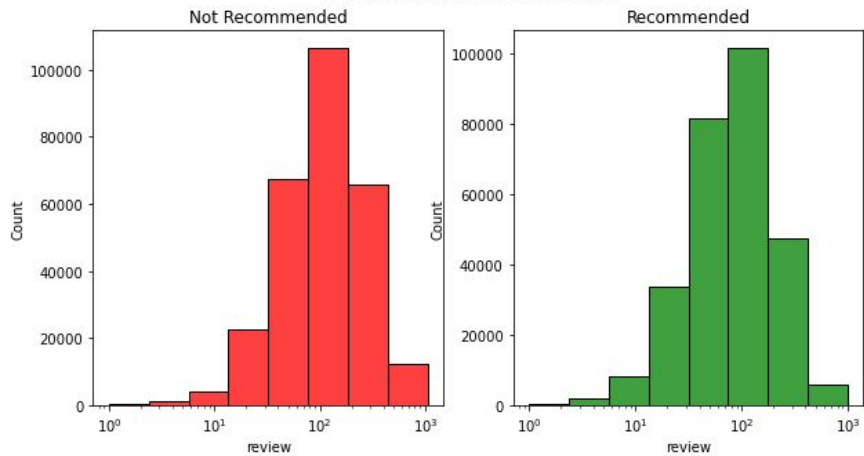
Restaurant Train

```
1 print('Recommended reviews: ', len(rest_train[rest_train['target']==1]))  
2 print('Not recommended reviews: ', len(rest_train[rest_train['target']==0]))
```

Recommended reviews: 280000

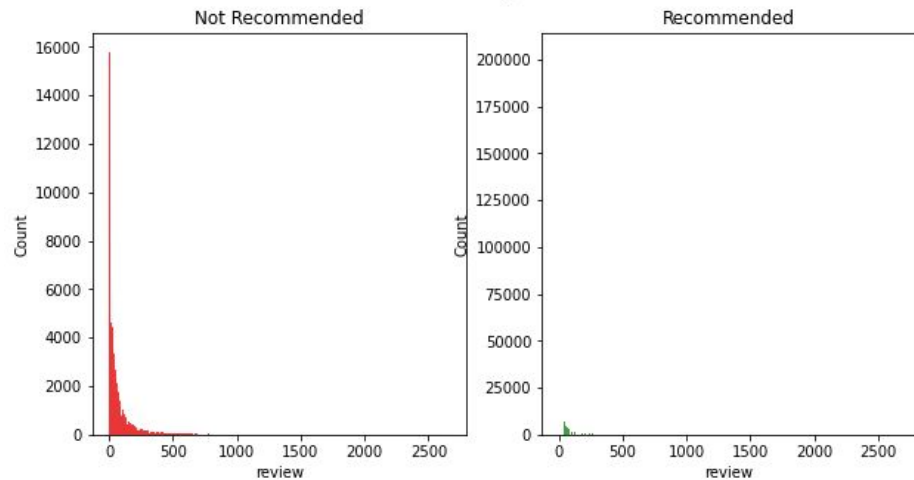
Not recommended reviews: 280000

Word amount in restuarant reviews



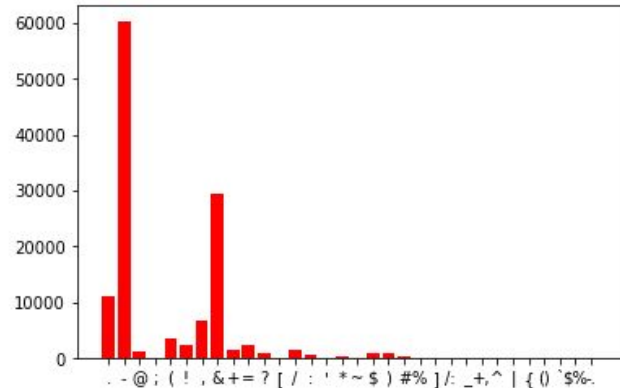
Log scaled

Word amount in game reviews

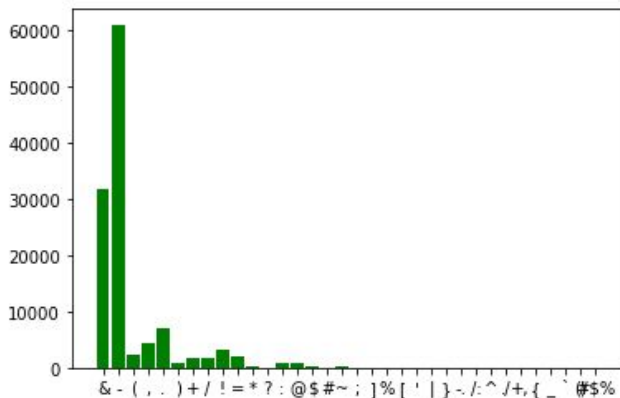


Unable to log scale due to error

Not Recommended

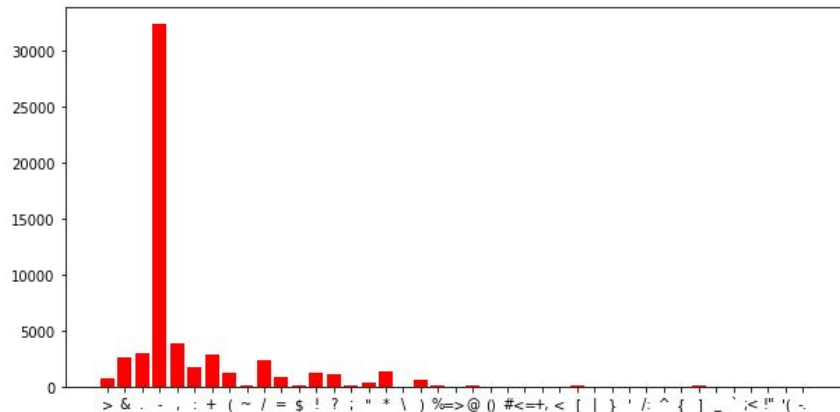


Recommended



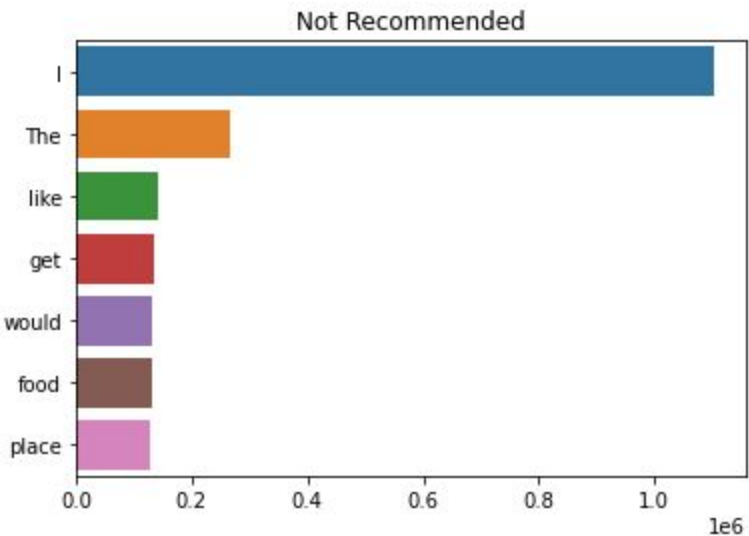
Restaurant Dataset
Punctuation Count

Not Recommended

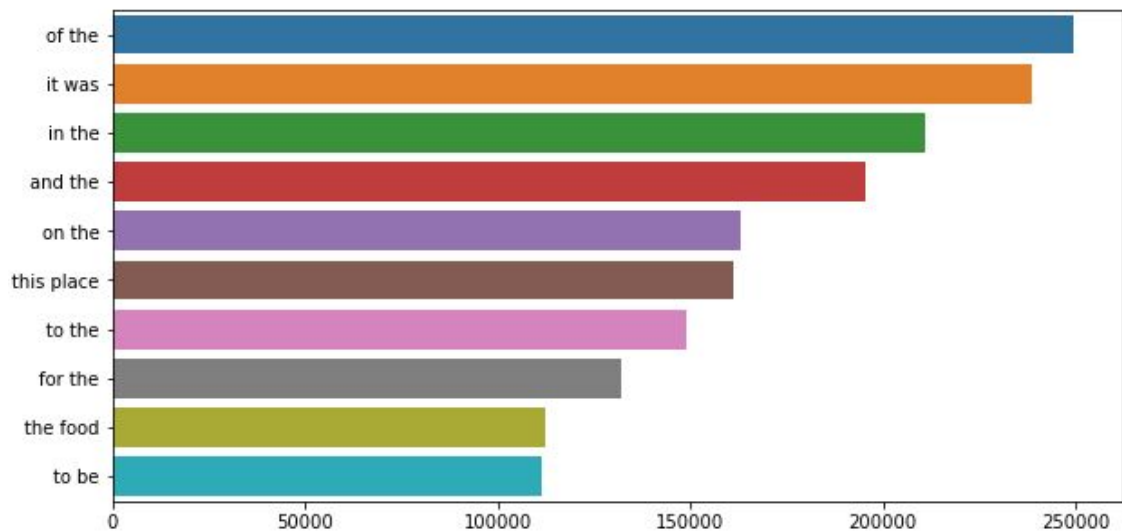


Game Dataset
Punctuation Count

Recommended could not
be graphed due to memory
error

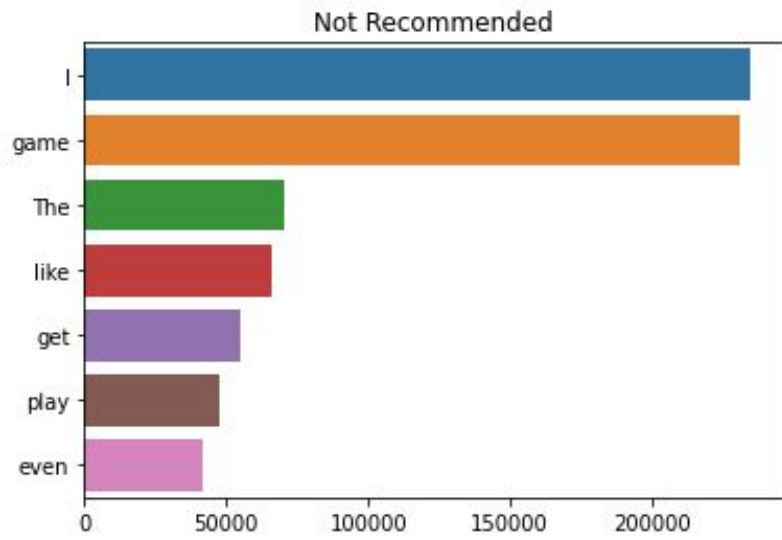


Common Word

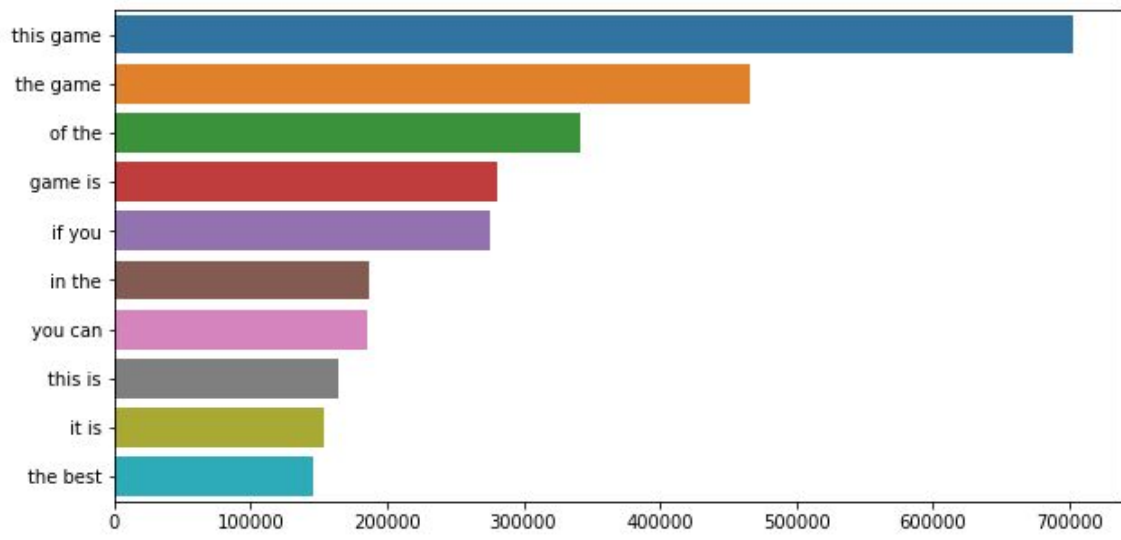


Common Bigram
(Word Combination)

Restaurant Dataset



Common Word

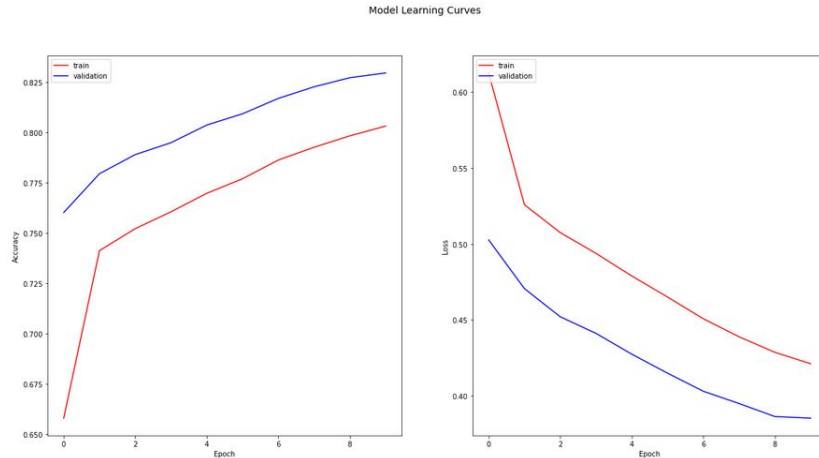


Common Bigram
(Word Combination)

Game Dataset

Model

- GLoVe Model
 - Deep learning Model
 - Epoch: 10
 - Batch size: 64
 - Loss: Binary Crossentropy
 - Optimizer: Keras Adam
 - Metrics: Accuracy



- Bert Model
 - Transformers Encoder model
 - Epoch: 10
 - Batch size: 16
 - Loss: Binary Crossentropy
 - Optimizer: Keras Adam
 - Metrics: Accuracy