CLASSIFICATION ANALYSIS: Hulu

Use Case #4: Competitive Intelligence

Team Members: Rijul Banerjee, Martha Lin, Adam Toy, Kathleen Sebastian, Supriya Shahane, Tatiksha Singh

OBJECTIVE

- Provide a **Hulu Product Manager** with an overview of **customer insights** on Hulu's performance in the marketplace

- This enables us to create **comparative benchmarks** that will be leveraged to build a competitive intelligence **chatbot and dashboard**

Methodology

01
Acquire Twitter Data
Utilize Twitter API to scrape tweets mentioning #Hulu

02
Tag Tweets
Create categories and tag tweets based on the categories. Divide tagged tweets into training and test data

03
Train Classification Model
Input training data into IBM Watson Assistant

04
Model Testing and Performance
Test Watson with the testing data and check model performance

05
Derive Insights
Key Learnings from the analysis

01 Acquire Twitter Data

Getting 1000 Hulu Tweets

```python
search_words = "#hulu"
date_since ="2019-10-01"

# Collect tweets
tweets = tw.Cursor(api.search,
                   q=search_words,
                   lang="en",
                   since=date_since).items(1000)

# Collect a list of tweets
[tweet.text for tweet in tweets]
```

```
['In honor of #Booksmart being on Hulu check out my review for it if you haven't already ! #Hulu #booksmartmovie… https://t.
co/sJwYe6131N',
 "RT @SuperReactBros: It's time for a little bit of revenge in our (and @NicholasMSlade's) reaction to the Official Trailer
for Reprisal.\n\n#R…",
 'RT @cacram81: Sign up for Disney+ through Rakuten\u200b &amp; get:\n$10 Cash Back on Disney+ Annual Subscription\n$7.50 Ca
```

Removing retweets

```python
new_search = search_words + " -filter:retweets"
new_search
tweets = tw.Cursor(api.search,
                   q=new_search,
                   lang="en",
                   since=date_since).items(1000)

tweet_loc=[tweet.text for tweet in tweets]
```

Storing tweets in a dataframe

```python
tweet_text = pd.DataFrame(data=tweet_loc,
                  columns=['tweet'])
tweet_text
```

	tweet
0	I asked #Alexa to "Play South Park on #Hulu" f...
1	There are only a few shows that I used to LOVE...
2	In honor of #Booksmart being on Hulu check out...
3	#DisneyPlus #Disney #Hulu #StarWars #Pixar #th...
4	No matter how broke we are.. We have enough mo...
...	...
995	Day Off in Bed Binge Watching @MaskedSingerFOX...
996	@hobybrenner @hulu @TabloTV @Philo Do Tablo an...
997	#CordCuttersNews \n\n@hulu Live tv subscribers...
998	@RememberEbola Visit my website ; https://t.co...
999	@Just_Zoom_In Visit my website ; https://t.co/...

1000 rows × 1 columns

HULU Tweets Word Cloud

02 Tag Tweets - 1

How did we tag tweets?

- Research papers for text classification

- Looked at Hulu Community Forums

- Group Consensus

Tweets	label 1	label 2	label 3	label 4
Can Hulu Bring Back East Los High? . . . Just Asking For A Friend . . Oh Screw It, I'm Asking For Myself. Bring It	organization	review	question	
Hulu (@hulu) Original WOUNDS (2019) Movie Review: Pure Nightmare Fuel By: @twistedtori87 #Horror #Hulu #Wounds	organization	entertainment	review	movie
Just signed up for #hulu and they are already upping their fees. So much for that.	Customer Service	organization	finance	account
Just joined #Hulu and they're already increasing their fees. Geesh!	Customer Service	organization	account	finance
Can we all agree that after watching this new episode of my hero Academia. Kirishima has become one of the top favorite	Review	tv_show	review	question
It's time for a little bit of revenge in our (and @NicholasMSlade's) reaction to the Official Trailer for Reprisal	Announcement	tv_show	review	
is #HULU worth getting? thinking of subscribing...	Product Question	organization	account	
Really @hulu ??? You raising the price $10 more starting in December?? Second time this year... this is getting ridiculous	Complaint	organization	account	
Episode two of our Castle Rock Review will premiere tomorrow! Pretty sure we are going to drop episode three the day	tv_show	new_release	review	

02 Tag Tweets – 2

How did we decide on the final categories?

Brainstormed on

- Granularity: kept it high

- Number: kept it minimum

- Sentiment: didn't add any positive or negative connotation to the categories as we were not looking to do any sentiment analysis at this stage

Tweets	Final Label
In honor of #Booksmart being on Hulu check out my review for it if you haven't already ! #Hulu #booksmartmovie	Review
#DisneyPlus #Disney #Hulu #StarWars #Pixar #theSimpsons #NatGeo #HBOGo #Marvel #streaming #streamingservice	Miscellaneous
No matter how broke we are.. We have enough money for #Netflix and #Hulu	Review
#Booksmart Is Just So Good. MOLLY AND AMYS FRIENDSHIP GIVES ME ALL THE FEELS So Glad It's On #Hulu!!!!!	Review
@notthetruelodge I NEED to know what's behind the door! #SaveLodge49 #amazon #hulu #appletv #netflix	Content
Can I openly say Iâ€™m loving Dollface without using the guilty pleasure a bit? Kat Dennings, the cameos, and do I	Review
Can Hulu Bring Back East Los High? . . . Just Asking For A Friend . . Oh Screw It, I'm Asking For Myself. Bring It	Content
I consider myself a prolific binger of movies and television. Possibly the greatest of our time. #bingewatching	Miscellaneous
Dollface is funny and worth watching #hulu #dollface	Review
Cord cut as of today! #youtubetv #netflix #hulu #DisneyPlus #AmazonPrimeVideo	Miscellaneous
Episode two of our Castle Rock Review will premiere tomorrow! Pretty sure we are going to drop episode three	Announcement
#hulu can't connect your account to #disneyplus you're better off cancelling #hulu completely and starting a new account	Customer Service

03 Train Classification Model

How did we train IBM Watson Assistant?

Prepared the training data

- Tagged 120 tweets

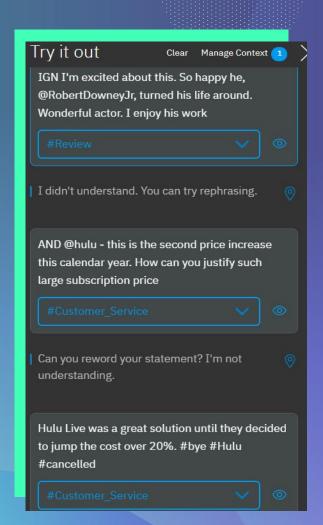- Cleaned unnecessary characters from the tweets

Trained the model

- Used import intents to feed our training data into IBM Watson Assistant

	Intents (5) ▲	Description	Modified ▲	Examples ▼
☐	#Announcement		an hour ago	31
☐	#Content		2 minutes ago	13
☐	#Customer_Service		an hour ago	23
☐	#Miscellaneous		a few seconds ago	23
☐	#Review		43 minutes ago	39

Final categories imported as Intents

04 Model Testing

- Testing data of 26 tweets
- Manually tagged the testing data with the actual label
- Tested the tweets in IBM Assistant
- Recorded the model output as Predicted Label

Test Tweets	Actual label	Predicted label	Model worked?
Only a few hours left and then it's to weekend baby !!! Getting some self care in this weekend. What are you doing	Miscellaneous	Miscellaneous	yes
Dang! I only got half FrankenDrew monologue #GH #hulu	Review	Announcement	no
Get smart on #streaming: the cost of unwatched content is NEGATIVE not zero. Good programming is "mostly a judgement"	Review	Review	yes
haven't even had hulu live tv for a month and they're raising the price like that. im sooo pissed #hulu @hulu	Customer_Service	Customer_Service	yes
CITY OF GOLD now available to see on #iTunes #GoolgePlay #Hulu #Amazon #AmazonUK #CityOfGold #IndieFilm #MadeOnMaui	Announcement	Announcement	yes
Hey @hulu @hulu_support I see you jacking up your live prices again. I also see you not jacking up the quality...	Customer_Service	Customer_Service	yes
How did I spend my Saturday? Oh you know just binge watching @DollfaceOnHulu #Dollface #hulu	Content	Review	no
Retweet if Over the Hedge should be put on #Netflix or #Hulu	Content	Content	yes
IGN I'm excited about this. So happy he, @RobertDowneyJr, turned his life around. Wonderful actor. I enjoy his wor	Review	Announcement	no

04 Model Performance

Intents	# of Test Tweets	Predicted correctly	Predicted incorrectly
Review	7	5	2
Customer Service	5	5	0
Content	7	1	6
Announcement	4	4	0
Miscellaneous	3	1	2
	26	16	10

Intents		Prediction				
		Review	Content	Announcement	Customer Service	Miscellaneous
Actual	Review	5		2		
	Content	3	1	1	2	
	Announcement			4		
	Customer Service				5	
	Miscellaneous	1			1	1

- Overall model performance = **60% accuracy**
- Model worked **100% for 2 intents** – customer service, announcement
- **Poor performance** for 2 intents – content, miscellaneous

Model Improvements

- Ran model on larger dataset
- Analyzed confusion matrix to find out where the prediction is incorrect
- Update the model based on revised categories

05 Key Insights

Need to create more specificity among tweets classification.
- Most tweets either fell under the 'review' or 'customer service' label
- Watson Assistant accurately predicted Announcement and Customer Service labels
  - Unsuccessful at predicting Content-related tweets
  - Reason: Some overlap with announcement and review-related tweets

Main keywords customers are using to talk about Hulu:
- Disney Plus, Dollface, Netflix, Price increase, Hulu support → brand perception
- Unjustified Price increase = neg brand perception

Moving Forward

Conduct Sentiment Analysis
- Customer service-related tweets & user reviews of new Hulu movies and shows
- Gain valuable insight to users & how they perceive Hulu's platform and new content

Integrate classification capability into chatbot and dashboard
- Enables Product Manager to quickly identify, classify and capture user's topic of discussion in the marketplace
- Expedites the decision-making process for PM

THANKS!

Does anyone have any questions?

Appendix

1. https://pdfs.semanticscholar.org/b396/2c869ea28161d77225914aab4a066aa016bf.pdf

2. https://medium.com/@urban_institute/how-to-categorize-large-amounts-of-text-data-71e1fd590e2a

3. http://cucis.ece.northwestern.edu/publications/pdf/LeePal11.pdf