A Sentiment Analysis of COVID-19 Tweets and its effects on Likeability and Retweet-ability

Andy Lee, supervised by Dr. Bilgehan Erdem
Student ID: 500163559

Ryerson University

Presentation Overview

1. Research
2. Overview
3. Pre-Python
4. Import
5. Cleaning
6. New Features
7. Basic Analysis
8. Linear Regression
9. Polynomial Regression
10. Comparison
11. Conclusion

Research Problem

To study whether a tweet's sentimentality and other features have an influence on the tweet's retweet-count and liked-counts.

Key Objectives

- Reduce Tweet text noises in our dataset
- Create new features using sentiment analysis tools to score each tweet
- Build different regression models
- Compare models metrics
- See whether it can predict "retweets" and "likes" based on independent features

Ryerson University

Lecture Review #1

- Sicilia et al. in “Twitter rumour detection in the health domain”
  - A detection system detecting rumours or non-rumour tweets in Health sector
  - Uses machine learning techniques: SVN, Nearest Neighbour, Random Forest

- Ravi in “A survey on opinion mining and sentiment analysis: Tasks, approaches and applications”
  - Comprehensive review and comparison of sentiment analysis techniques from over 300 papers
  - Result shows room for growth in intelligence-based techniques such as Random Forest

Lecture Review #2

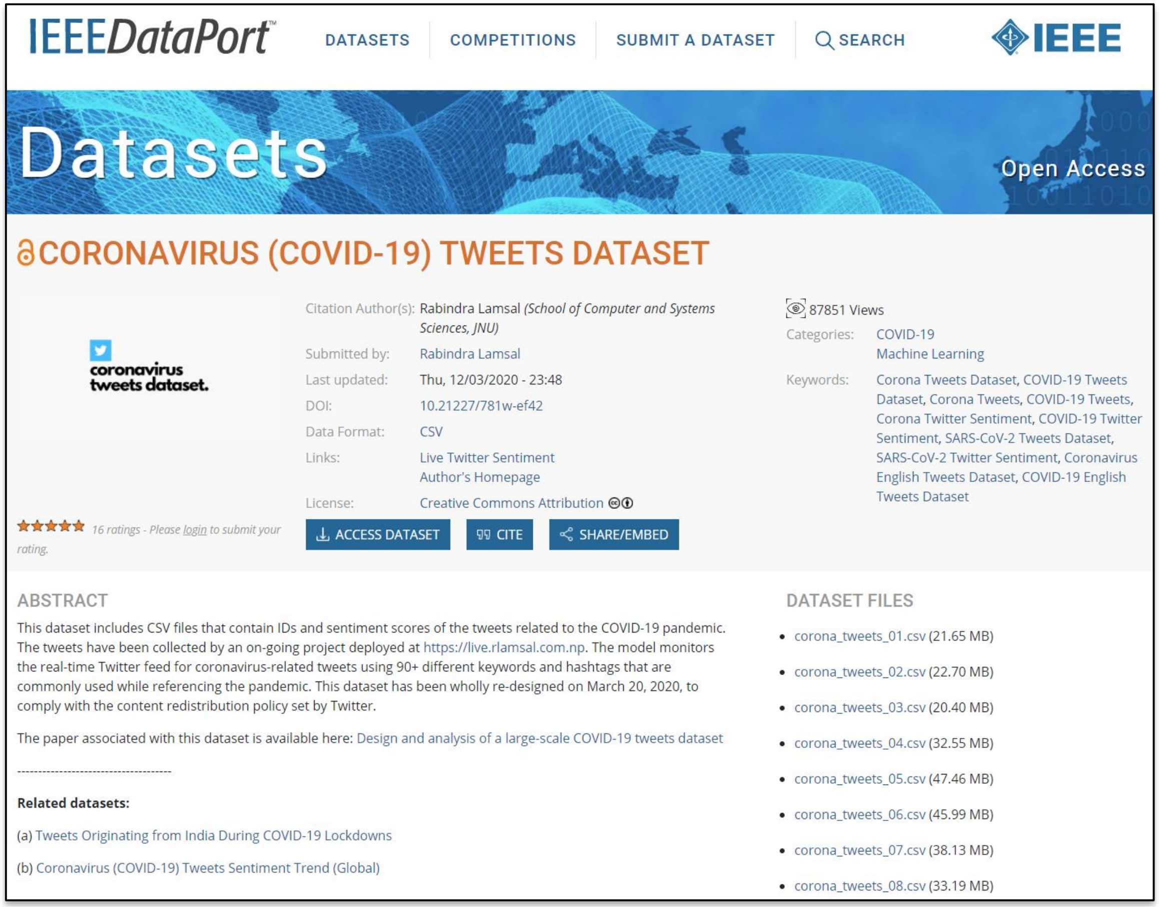
- Carlos et al. in "Detecting and Monitoring Hate Speech in Twitter"
  - A detection system using text and emoji to detect hate speech
  - Experimented 19 different strategies of feature and classification models
  - Concluded combining LTSM and MLP-NN produces the best result
- Zubiaga et al. in "Detection and Resolution of Rumours in Social Media: A Survey"
  - Describes a rumour detection system combining: rumour detection, tracking, stance classification, and veracity classification
- In our study
  - Our dataset is relatively new (March 2020) from IEEE
  - A combination of NLP and Regression
  - Other studies uses more sophisticated tools and techniques to tackle the problem

Overview of Project Approach

1. Download dataset from IEEE
2. Load dataset onto Hydrator program convert Tweet ID to Tweets
3. Extracted tweets is imported to Pandas
4. Perform data cleaning
5. Perform sentiment analysis on tweets to introduce new features
6. Perform basic analysis on the dataset
7. Build prediction model with linear regression, then with *k*-fold
8. Build prediction model with polynomial regression, then with *k*-fold
9. Compare the results

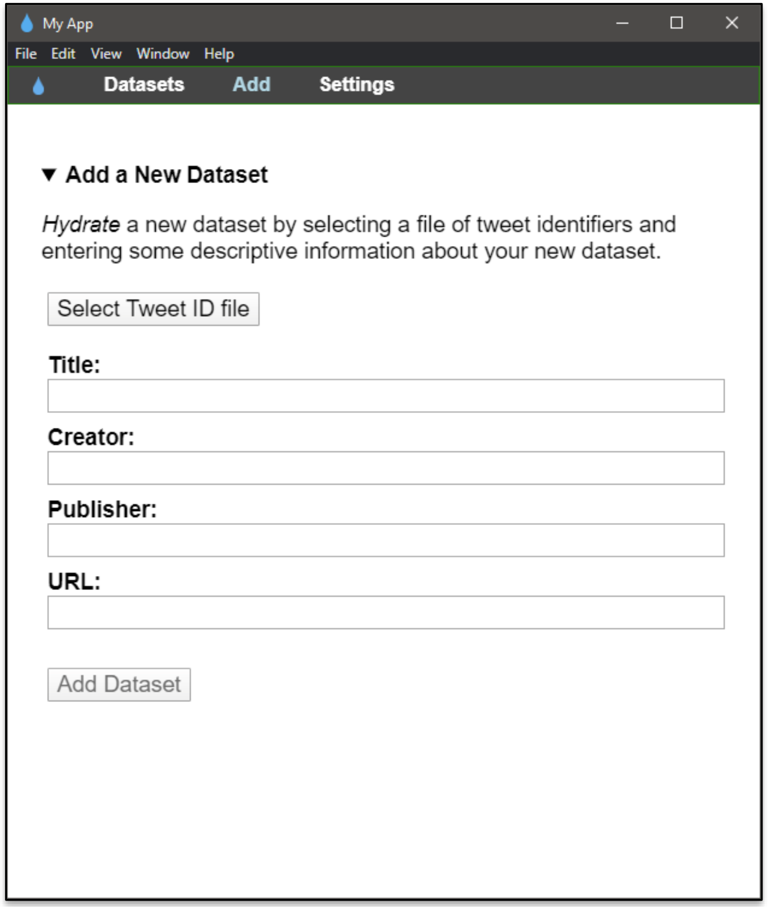
(Pre-Python) Download dataset

- Dataset from IEEE
  - https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset
  - Since March 20, 2020
  - Updated everyday
  - Filtered 90+ COVID-19 related keywords and hashtags (https://rlamsal.com.np/keywords.tsv)
- This project uses dataset #13
  - Tweets March 31-April 1, 2020



(Pre-Python) Hydrator

- Dataset loaded to Hydrator program
  - Convert Tweet ID to Tweets using Twitter API
  - Does not need developer’s account
  - Limitations per 15 mins
  - Takes few hours
  - Returns multidimensional JSON
    - flatten to CSV through Hydrator
- Convert .csv dataset
  - Large in filesize
  - Can use external programs to trim features



Features Overview

Twitter API returns 35 features, and we will be using 8 features:

Feature	Description
favorite_count	No. of "liked" this tweet has
retweet_count	No. of retweet this tweet has
user_followers_count	No. of follower Tweet's account has
user_statuses_count	No. of tweets and retweets this user issued
user_favourites_count	No. of tweets this user liked
user_followers_count	No. of followers this account has
user_friends_count	No. of users this account follows
user_listed_count	No. of public lists this user is a member of
Compound	the overall sentiment score calculated; -1 is negative, 0 is neutral, 1 is positive

Dependent Variables

New added feature

Reference: Twitter Developer Documentation. (2020). Retrieve from: https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview/user-object

Data Cleaning, Manipulating, and Guessing Language

- Data cleaning such as
  - Converting features to the correct data type
  - Handling null values
- Reduce noise in tweets
  - Removing \n, URL, user referrals, and hashtags

	text
714006	RT @Maashish81us: Live #CoronaJihad\n\nIn india , these terrorist skull caps are spreading Corona in india\n\n80% Corona Infected people are #M…
368670	Good.\nPlease always try to tweet about highlighting positive things of life. \n\nNo use going after IK/PTI.\n\nYou can and should contribute alot more in different ways.\n\n@RehamKhan1 https://t.co/IkvY8XibLF

- Parse each tweet to filter out observations that are non-English (fully/partially)
  - Of the 245K tweets, 94.9% was detected as English
  - The remainder is excluded from further study

501093	@voiceaditya @ANI Infected mulla ko usi masjid me maulvi ke sath band kiya jaye. \nNo need to give them any medical facilities. Let them to die by theirs itself spread corona. \nThey are doing terror activities.

Adding New Features with Sentiment Analysis

- NLTK library, SentimentIntensityAnalyzer class, polarity_score() method
- Create new features by running polarity_score() on each tweet
  - Returns a dictionary object positive, neutral, negative, compound
- Compound score is normalization of the sum of valence
  - norm_score = score/math.sqrt((score*score) + alpha)
  - "score" is computed based on some heuristics, sentiment lexicon
  - The normalized score is between -1 and 1

	modified_text	sentimentscore
714555	Why is in Australia airing the TRUTH about China covering up the Corona Virus and wont air it here?	{'neg': 0.0, 'neu': 0.856, 'pos': 0.144, 'compound': 0.4648}
174370	Corona Virus Live	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

Reference: https://stackoverflow.com/questions/40325980/how-is-the-vader-compound-polarity-score-calculated-in-python-nltk

Observations from Tweets and Compound Score

- Expand the dictionary object to 4 new features
- Focus on the “compound” score feature
- Observe Tweet text and its compound score

	modified_text	compound
615866	Our World after Corona	0.0000
706139	Congratulations to all over the and specially to Pakistani People 😇😇😇😇	0.5994
494723	This 👎👎👎	0.0000
444492	... the IOF are filth, Hazmat suits don't contain their scum.	0.0000
186291	Corona made you old Wes 😂😂	0.0000
533428	The American Taliban strikes again.	-0.3612
338328	50 New positive cases in TodayTotal number of positive cases in stands at 124.	0.8176
426248	Glad to hear talking constructively about how Israel and PA officials need to deepen existing cooperation as we are living in one interconnected space. Need to save people, Jews and Arabs, Israelis and Palestinians from this awful virus.	0.4019
211681	Na inside Corona season u see person wey dey knock u off your own feet.	0.0000
104640	To all the doctors, nurses, police men, service men, honourable Prime Minister & every Indian fighting Corona, I believe in the power of social media. We stand strong with you'll. We'll stay home to save India. A poem written by me!	0.8122

Minimal to no feedback on sarcasm

“Positive” most likely brought the values up

Does a better job at scoring proper sentences

	modified_text	compound
672670	Stay safe America @ San Antonio, Texas	0.4404
518154	How the numbers are reported by the Media to paint the Pandemic Drama!	0.0000
263674	On aajtak with says - On 28th March tableegi leader given speech "...corona lock down is nothing but a sajish to keep muslim away from mosques. There is no better place to die other than a mosque" Is this true ?	0.0194
678751	Infotainment News TV: Coronavirus: France reports record 499 deaths in a...	0.0000
446913	Fear of corona	-0.4939
10781	Trump cannot remember that he learned of the corona virus in January!	0.0000
524974	times of corona be like	0.3612
619678	What else Chinese doing to medicine you take?	0.0000
366643	people are anot understanding the fear of corona and lockdown. They are still roaming outside the house. What the administration is doing to ensure of lockdown. This is the location of bank road, doma pokhar behind adarsh vidhya mandir school	-0.1531
170284	Maybe hate on Bill Gates instead of Trump, because ask yourself WHY he and other elites ran a “Corona virus simulation” back in October 2019! The media tries to refute this but it’s too late. It’s out of the bag. Their op went LIVE and tried to kill us. >>	-0.8805
583298	My Corona virus presentation So at home we are seeing who can do the best presentation we even put our videos on Facebook to, so here's mine that I created.	0.7351
726870	While the attached photo speaks of corona, it also shows how distracting the impeachment process was to Dear Leader.	0.1027
710616	omg wait BLACK MIRROR CORONA VIRUS EPISODE	0.0000
223301	It will end in corona 😔😔	0.0000
430473	I SAW THIS IN MY DREAMSWHAT DOES MEAN#TuesdayThoughts	0.0000
302884	Honorable prime minister of India good evening.Please give a chance to Dr Biswaroop Roy Chowdhury to cure corona patients 🙏	0.8126
46471	A New Year Has Been Added To The Chinese Astrology Calendar- Year of the Corona	0.0000
414246	You made a list of the succession of Trump and the Corona Virus. Where do I find it..thx	0.2023
487	You need to view this video, DeAnna! It will blow your mind!	0.0000
84323	Coronavirus: Record supermarket sales in March ‘busier than Christmas’	0.0000

Lack of context, hard to gauge

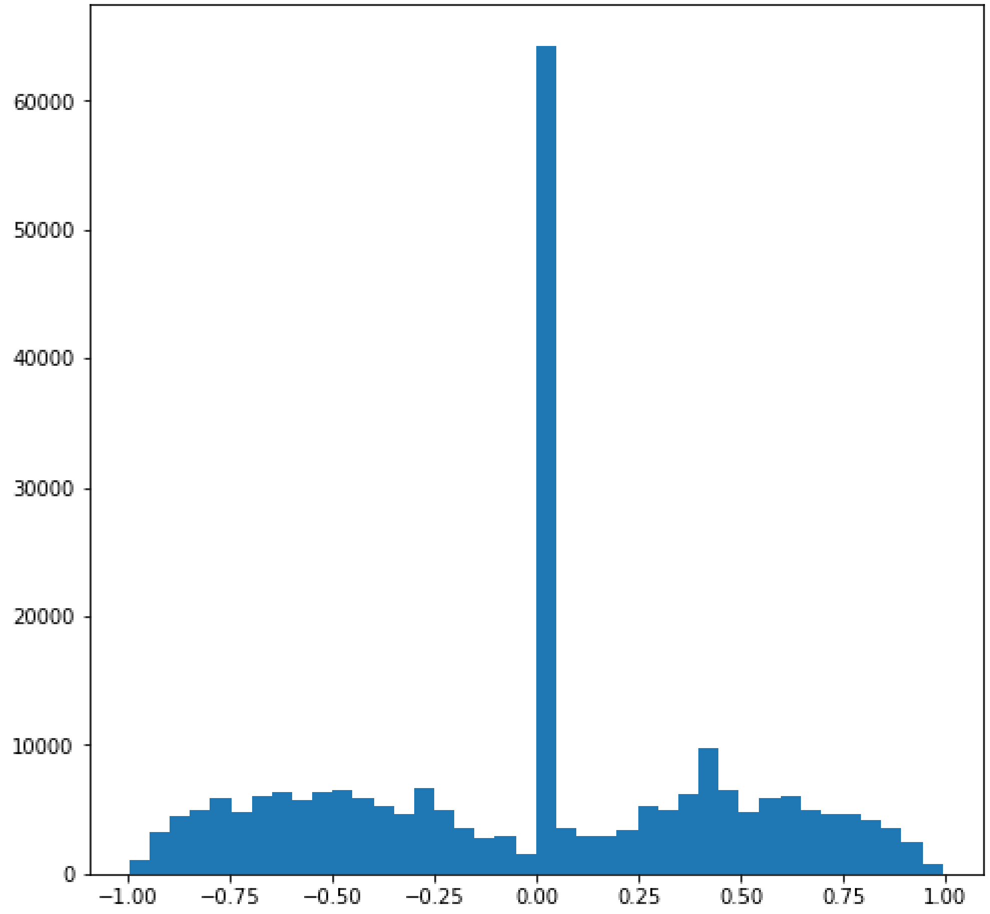
Better at scoring complete sentences

Minimal to no feedback on sarcasm

Ryerson University

Observations based on the Compound feature

- SentimentIntensityAnalyzer's polarity_score Summary
  - The Good
    - Does a good job when given complete sentences for both positive or negative
  - The Bad
    - Minimal to no feedback on sarcasm
      - A lot of 0.00 scores
      - Appears to give as neutral score on texts it could not analyze
    - Individual tweets lacks context – affect score
    - Some words such as "positive" can trick the algorithm

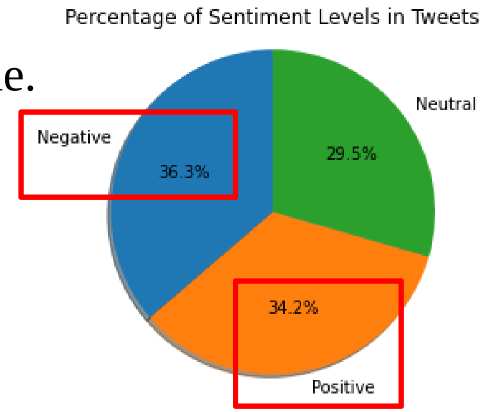


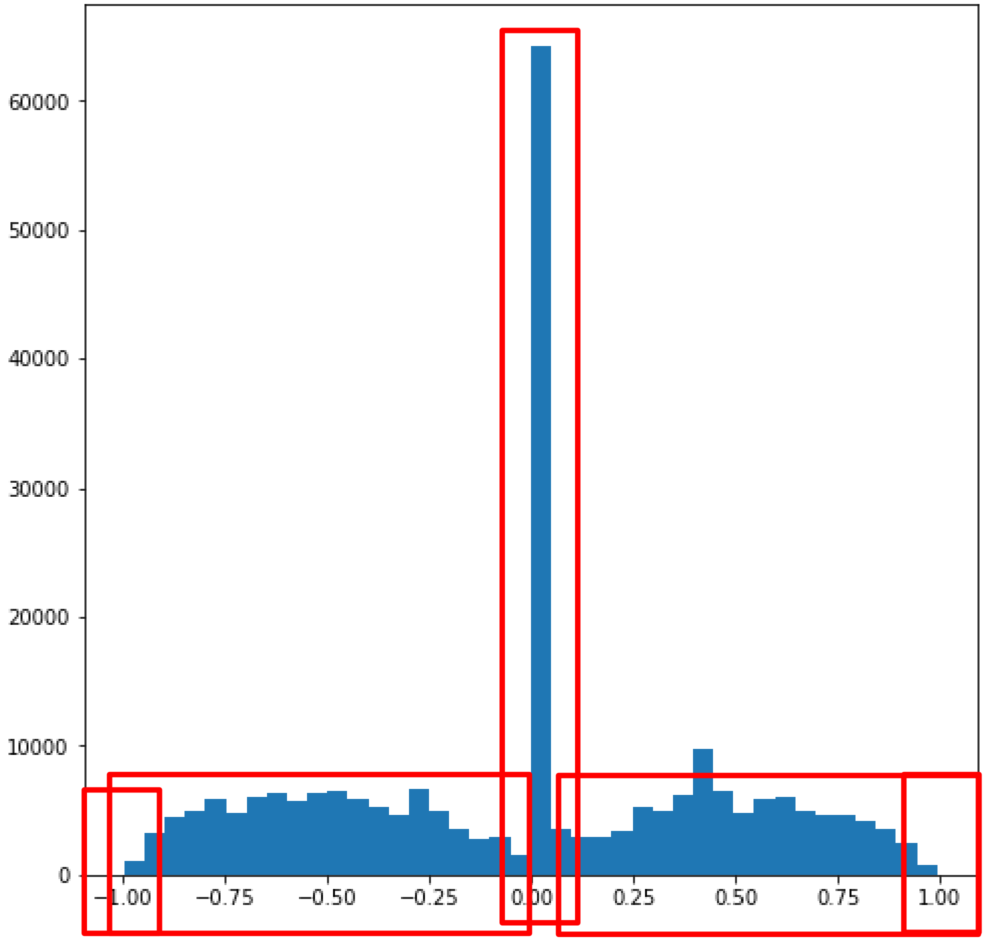
Distribution of the compound scores between -1 to 1 in 40 bins.

Observations based on the Compound feature

- Classification negative, neutral, positive bins based on Deo et al. research paper.
  - [-1 to -0.1][-0.1 to 0.1][0.1 to 1]
- Approx. a quarter observations in neutral range
- Extremes at -1.00 and +1.00 has the lowest count
- Multimodal
- It is approximately symmetric
  - Positive and negative are about the same.

```
neg     89051
pos     84039
neu     72299
```





Distribution of the compound scores between -1 to 1 in 40 bins.

Deo, G. S., Mishra, A., Jalaluddin, Z. M., & Mahamuni, C. V. (2020, September). Predictive Analysis of Resource Usage Data in Academic Libraries using the VADER Sentiment Algorithm. In 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 221-228). IEEE.

Basic Analysis – Features Correlation

- Favorite_count and user_followers_count have high correlation
- Compound has little correlation with other features

	favorite_count	retweet_count	user_followers_count	compound	user_statuses_count	user_favourites_count	user_friends_count	user_listed_count
favorite_count	1.000000	0.908576	0.165683	-0.001108	0.015366	0.007852	0.005575	0.051646
retweet_count	0.908576	1.000000	0.090025	-0.003071	0.015991	0.007261	0.007084	0.039251
user_followers_count	0.165683	0.090025	1.000000	0.003693	0.116452	0.000535	0.032721	0.609057
compound	-0.001108	-0.003071	0.003693	1.000000	-0.022604	-0.028941	-0.006560	0.000249
user_statuses_count	0.015366	0.015991	0.116452	-0.022604	1.000000	0.326723	0.105958	0.128835
user_favourites_count	0.007852	0.007261	0.000535	-0.028941	0.326723	1.000000	0.100920	0.016049
user_friends_count	0.005575	0.007084	0.032721	-0.006560	0.105958	0.100920	1.000000	0.048344
user_listed_count	0.051646	0.039251	0.609057	0.000249	0.128835	0.016049	0.048344	1.000000

Basic Analysis – Mean and Standard Deviation

- SD is relatively high which means the distribution is fairly spread out.

	favorite_count	retweet_count	user_followers_count	compound	user_statuses_count	user_favourites_count	user_friends_count	user_listed_count
count	245389.00000	245389.00000	245389.00000	245389.00000	245389.00000	245389.00000	245389.00000	245389.00000
mean	9.52971	2.70475	9251.20519	-0.01655	27468.30799	18028.51423	1284.43071	34.64758
std	384.21091	122.27732	250678.23602	0.47919	75560.36124	39592.12977	10409.72569	680.04372
min	0.00000	0.00000	0.00000	-0.99520	1.00000	0.00000	0.00000	0.00000
25%	0.00000	0.00000	64.00000	-0.39760	1156.00000	510.00000	135.00000	0.00000
50%	0.00000	0.00000	331.00000	0.00000	6344.00000	4200.00000	399.00000	1.00000
75%	1.00000	0.00000	1307.00000	0.36120	24616.00000	17807.00000	1038.00000	6.00000
max	108137.00000	35137.00000	62855265.00000	0.99190	4472178.00000	1254400.00000	4322723.00000	202433.00000

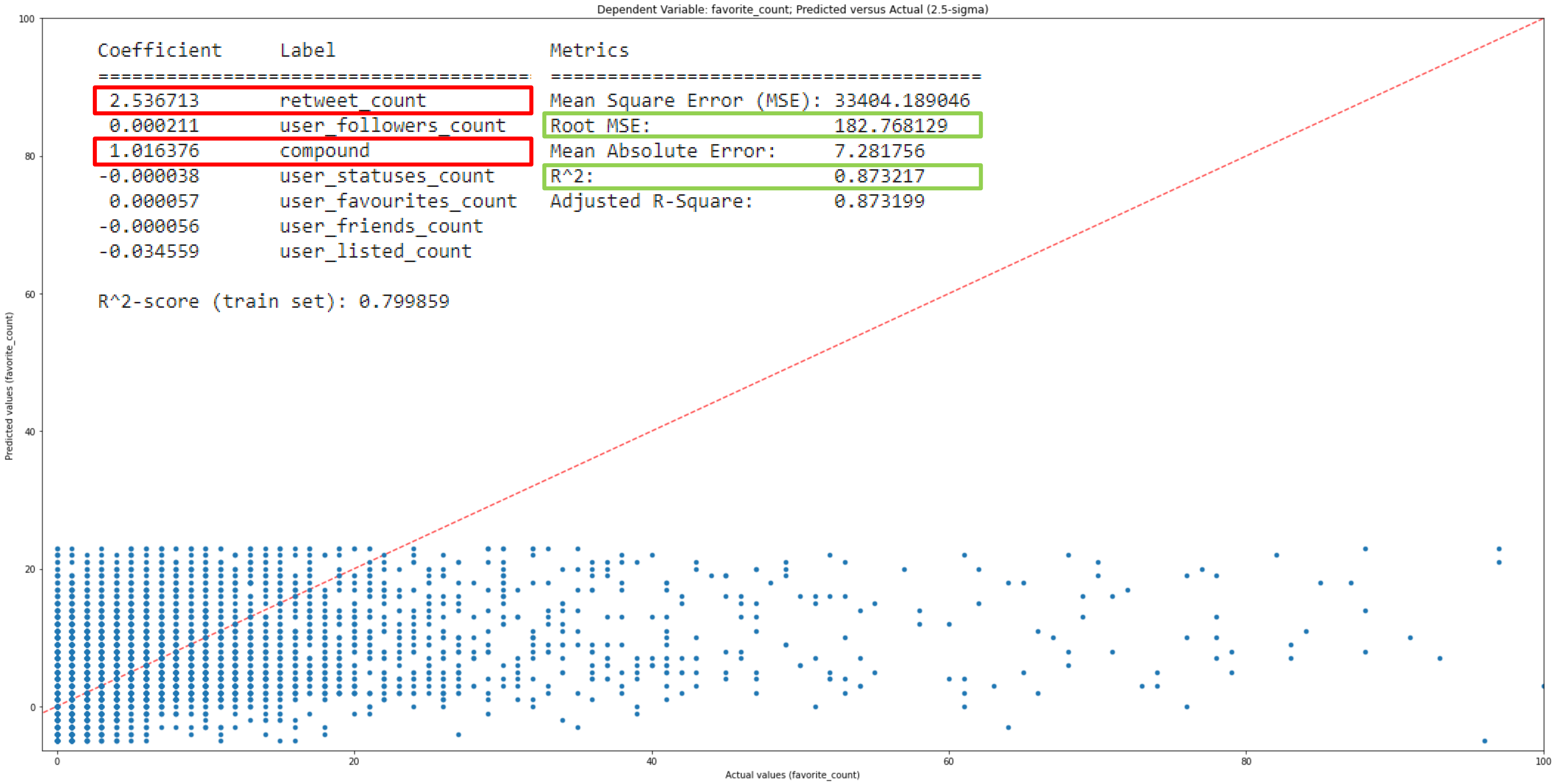
Building Different Regression Models

- Regression
  - Linear Regression 80% Train, 20% Test
  - Linear Regression in *k*-Fold (k=5)
  - Polynomial Regression, then *k*-Fold
- Two dependent variables

1. favorite_count
2. retweet_count

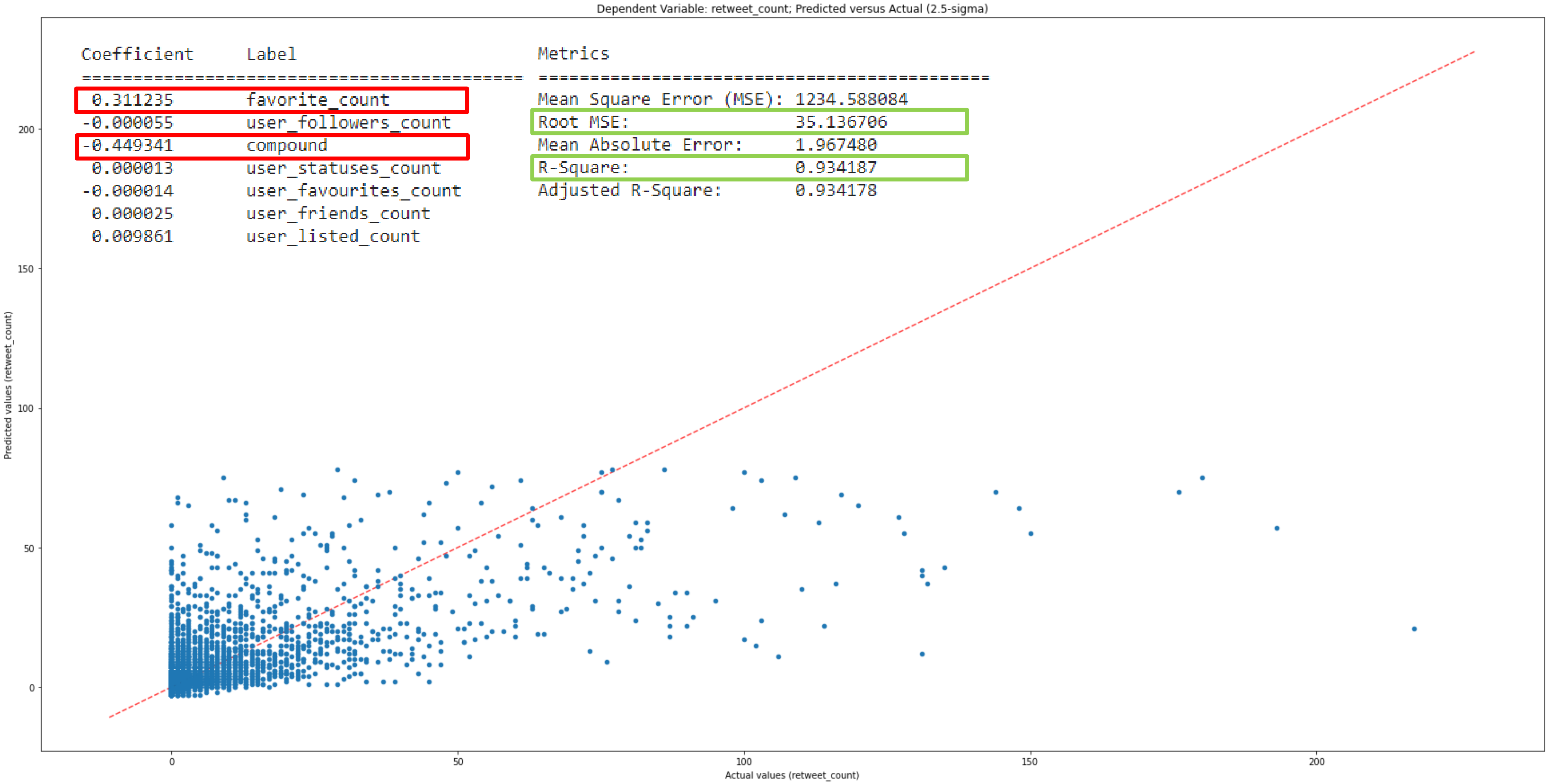
Feature	Description
favorite_count	No. of "liked" this tweet has
retweet_count	No. of retweet this tweet has
user_followers_count	No. of follower Tweet's account has
user_statuses_count	No. of tweets and retweets this user issued
user_favourites_count	No. of tweets this user liked
user_followers_count	No. of followers this account has
user_friends_count	No. of users this account follows
user_listed_count	No. of public lists this user is a member of
Compound	Sentiment score calculated between -1 to 1

Linear Regression – “Favourite Count” as Dependent



Ryerson University

Linear Regression – “Retweet Count” as Dependent



Linear Regression + K-fold

#1 **favorite_count**; Linear Regression + K-fold

```
Dependent Variable: favorite_count
Independent Variables: ['retweet_count', 'user_followers_count', 'compound', 'user_statuses_count', 'user_favourites_count', 'user_friends_count', 'user_listed_count']
K-fold: 5
MSE                     RMSE                    R-Square
==========================================================================================
31564.218170            177.663216              0.877825
10521.044828            102.572145              0.704872
130028.329157           360.594411              0.378130
10211.723622            101.053073              0.782832
13119.346895            114.539718              0.930201

Average
==========================================================================================
39088.932535            171.284512              0.734772
```

```
Metrics
========================================
Mean Square Error (MSE): 33404.189046
Root MSE:                182.768129
Mean Absolute Error:     7.281756
R^2:                     0.873217
Adjusted R-Square:       0.873199
```

#2 **retweet_count**; Linear Regression + K-fold

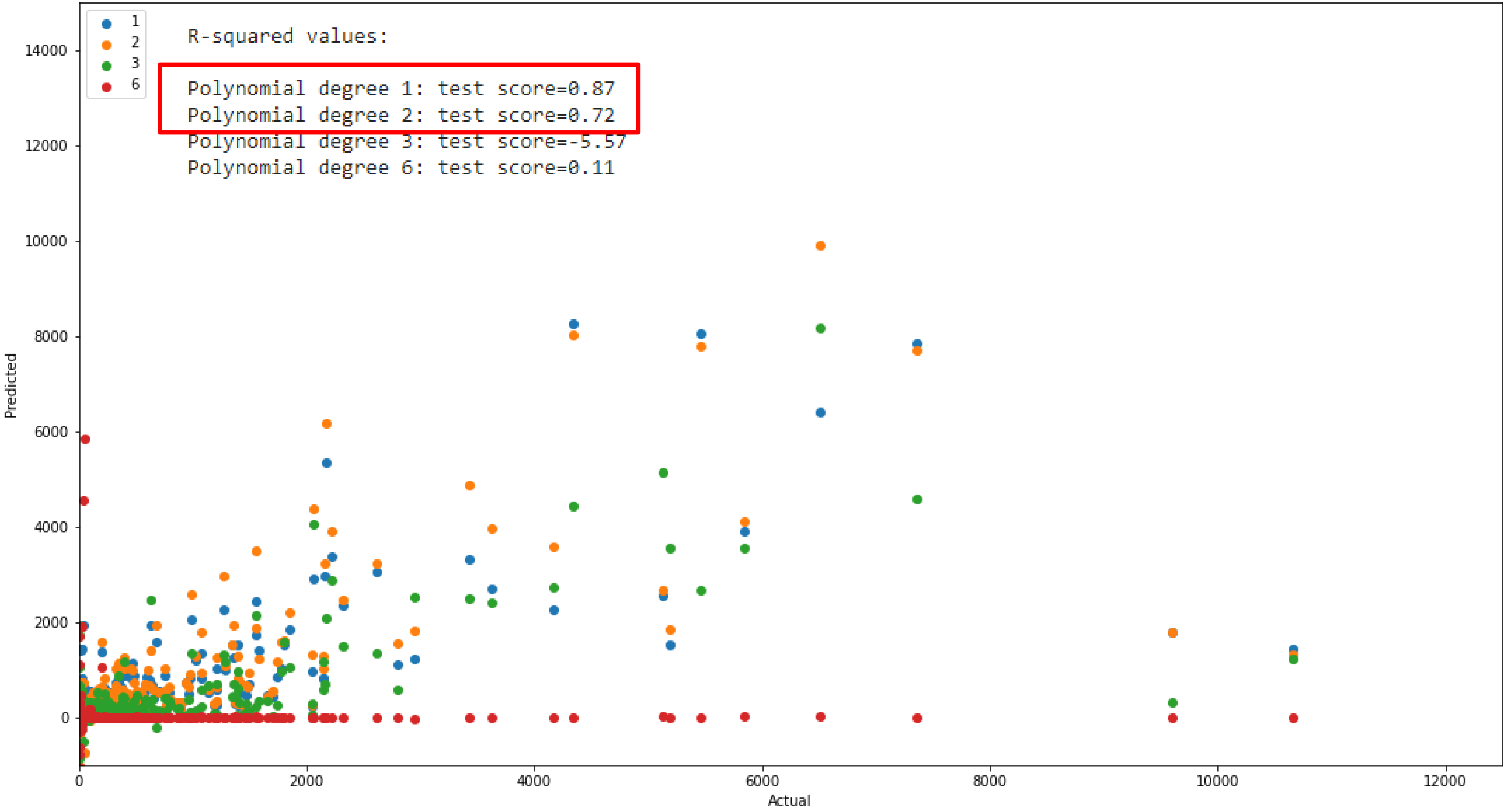
```
Dependent Variable: retweet_count
Independent Variables: ['favorite_count', 'user_followers_count', 'compound', 'user_statuses_count', 'user_favourites_count', 'user_friends_count', 'user_listed_count']
K-fold: 5
MSE                     RMSE                    R-Square
==========================================================================================
1001.568224             31.647563               0.945563
798.111384              28.250865               0.605363
10974.758804            104.760483              0.685319
728.033332              26.982093               0.750776
666.921016              25.824814               0.959674

Average
==========================================================================================
2833.878552             43.493164               0.789339
```

```
Metrics
==========================================
Mean Square Error (MSE): 1234.588084
Root MSE:                35.136706
Mean Absolute Error:     1.967480
R-Square:                0.934187
Adjusted R-Square:       0.934178
```

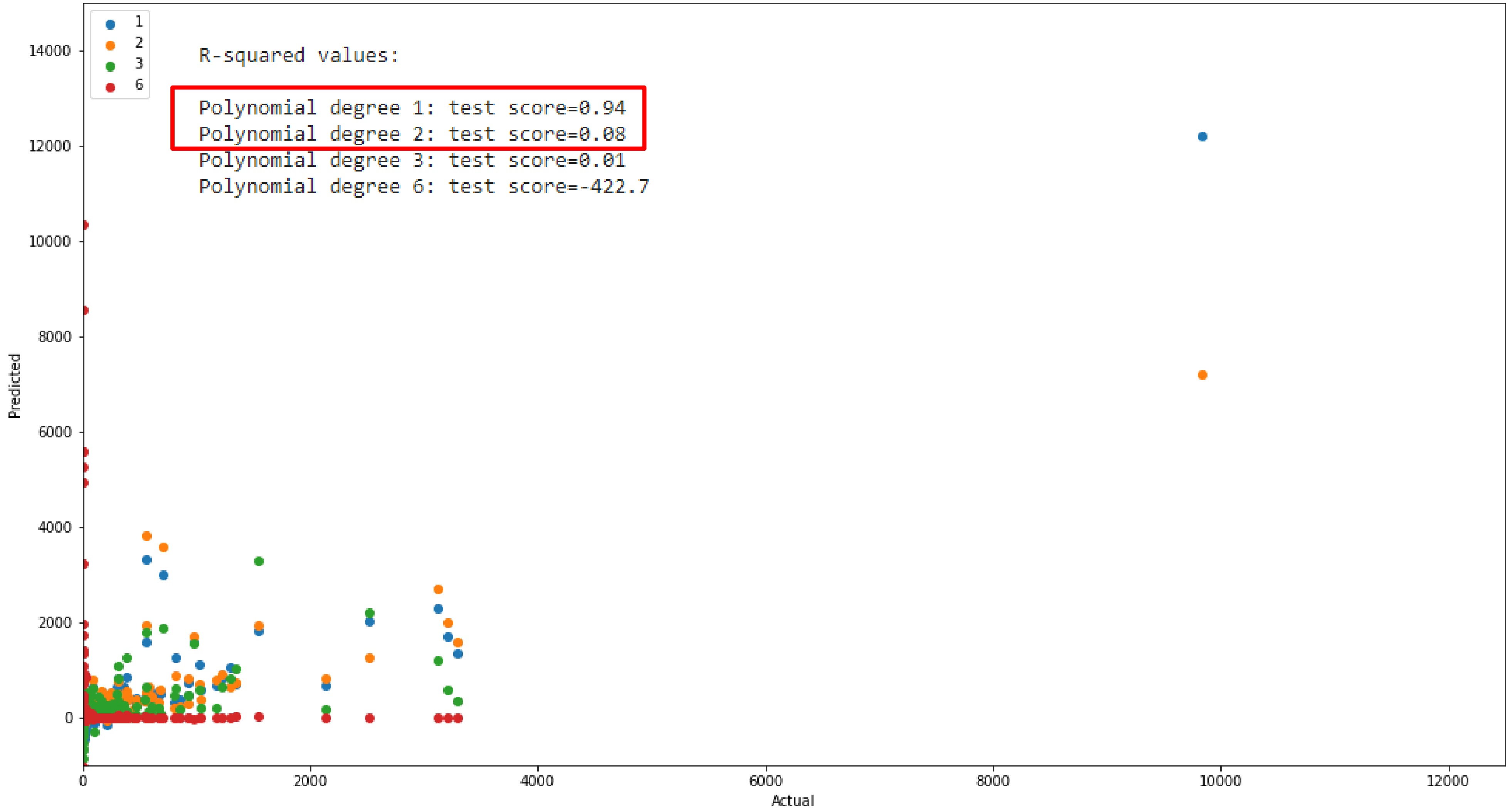
Ryerson University

Polynomial Regression – “Favorite Count” as Dependent



Ryerson University

Polynomial Regression – "Retweet Count" as Dependent



Ryerson University

Regression Comparison Results, “Favourite Count” as Dependent

Polynomial Regression + K-fold

```
Dependent Variable: favorite_count
Independent Variables: ['retweet_count', 'user_followers_count', 'compound', 'user_statuses_count', 'user_verified']
K-fold: 5
Polynomial               Root MSE                         R-Square
==============================================================================================
Degree 1:                176.1003309179579                0.88
Degree 2:                283.8671924780902                0.69
Degree 3:                906.2368606263769                -2.18
Degree 6:                856.3979925552468                -1.84
----------------------------------------------------------------------------------
Degree 1:                101.57117876732413               0.71
Degree 2:                113.61553501192776               0.64
Degree 3:                164.8107158447843                0.24
Degree 6:                1067.8173629126154               -30.98
----------------------------------------------------------------------------------
Degree 1:                361.7648735168686                0.37
Degree 2:                578.1225176563707                -0.6
Degree 3:                10089.810890024544               -485.89
Degree 6:                38011377.335343964               -6910168447.01
----------------------------------------------------------------------------------
Degree 1:                106.07747474685985               0.76
Degree 2:                132.28407898975604               0.63
Degree 3:                429.9725980224988                -2.93
Degree 6:                21654.500707596068               -9971.23
----------------------------------------------------------------------------------
Degree 1:                113.14836415856713               0.93
Degree 2:                284.245851156108                 0.57
Degree 3:                1548.2663280761865               -11.75
Degree 6:                15713.147572038102               -1312.61
----------------------------------------------------------------------------------
Degree   Average RMSE     Average R^2
1         171.732444       0.730000
2         278.427035       0.386000
3         2627.819479     -100.502000
6         7610133.839796  -1382035952.734000
```

Linear Regression + K-fold

```
Dependent Variable: favorite_count
Independent Variables: ['retweet_count', 'user_followers_count', 'compound', 'user_statuses_count', 'user_favourites_count', 'user_friends_count', 'user_listed_count'
K-fold: 5
MSE                     RMSE                    R-Square
==========================================================================================
31564.218170            177.663216              0.877825
10521.044828            102.572145              0.704872
130028.329157           360.594411              0.378130
10211.723622            101.053073              0.782832
13119.346895            114.539718              0.930201

Average
==========================================================================================
39088.932535            171.284512              0.734772
```

Regression Comparison Results, “Retweet Count” as Dependent

Polynomial Regression + K-fold

```
Dependent Variable: retweet_count
Independent Variables: ['favorite_count', 'user_followers_count', 'compound', 'user_statuses_count', 'user_verified']
K-fold: 5
Polynomial              Root MSE                        R-Square
=============================================================================================
Degree 1:               30.619157472439177              0.95
Degree 2:               143.57465611146827              -0.12
Degree 3:               197.92059429558387              -1.13
Degree 6:               140.91449714880363              -0.08
---------------------------------------------------------------------------------
Degree 1:               27.778555948274388              0.62
Degree 2:               33.17703268992806               0.46
Degree 3:               44.99600146557491               -0.0
Degree 6:               1595.006459641652               -1256.94
---------------------------------------------------------------------------------
Degree 1:               105.01778633851833              0.68
Degree 2:               109.10569803537199              0.66
Degree 3:               1123.4793102354063              -35.19
Degree 6:               185441.21308761396              -986024.03
---------------------------------------------------------------------------------
Degree 1:               28.269222892920038              0.73
Degree 2:               47.405288929894574              0.23
Degree 3:               55.005669883911004              -0.04
Degree 6:               10069.950456869094              -34712.13
---------------------------------------------------------------------------------
Degree 1:               25.580491200820568              0.96
Degree 2:               73.46282149176497               0.67
Degree 3:               241.73438207411132              -2.53
Degree 6:               5081.564815280062               -1560.37
---------------------------------------------------------------------------------
Degree  Average RMSE    Average R^2
1        43.453043       0.788000
2        81.345099       0.380000
3        332.627192     -7.778000
6        40465.729863   -204710.710000
```

Linear Regression + K-fold

```
Dependent Variable: retweet_count
Independent Variables: ['favorite_count', 'user_followers_count', 'compound', 'user_statuses_count', 'user_favourites_count', 'user_followers_count', 'user_fri
K-fold: 5
MSE                     RMSE                    R-Square
=============================================================================================
1000.984587             31.638340               0.945595
798.111384              28.250865               0.605363
10974.758804            104.760483              0.685319
726.632995              26.956131               0.751255
668.976723              25.864584               0.959550

Average
=============================================================================================
2833.892899             43.494081               0.789416
```

Conclusion

- Regression models were somewhat able to predict Retweet-ability (Retweet count) and Likeability (Favorite_count)
- Linear Regression produced slightly better results, at less resource cost
- Compound feature contributed to the prediction
- We can say that "a tweet's sentimentality and other features have some influence on the tweet's retweet-count and liked-counts"

Further Studies

- Use classification instead of regression approach
  - Use categorical instead of continuous value for sentiment score
- Conduct research on the 3 sentiment groups (negative, neutral, positive) independently
  - The 3 groups might have its own regression fit
- Group by user's screen name and calculate each user's mean sentiment score
  - Research whether there are any effects on number of followers and number of favourite tweets against a user's mean sentiment score

References #1

- Bode, L., & Vraga, E. K. (2017). See Something, Say Something: Correction of Global Health Misinformation on Social Media. *Health Communication, 33*(9), 1131-1140. doi:10.1080/10410236.2017.1331312
- Caplan, L. (2017, 10 11). *Should Facebook and Twitter Be Regulated Under the First Amendment?* Retrieved from WIRED: https://www.wired.com/story/should-facebook-and-twitter-be-regulated-under-the-first-amendment/
- Carvalho, L. d., Silva, M. V., Costa, T. d., Oliveira, T. E., & Oliveira, G. A. (2020, 9). Public Health and the COVID-19 pandemic: challenges for global health. *Research, Society and Development, 9*(7). doi:10.33448/rsd-v9i7.4188
- Culliford, E. (2020, 10 6). *Twitter is testing how its misinformation labels can be more obvious, direct*. Retrieved from Reuters: https://ca.reuters.com/article/us-usa-election-twitter-labels-focus-idCAKBN26R1ML
- Deo, G. S., Mishra, A., Jalaluddin, Z. M., & Mahamuni, C. V. (2020). Predictive Analysis of Resource Usage Data in Academic Libraries using the VADER Sentiment Algorithm. *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, 221-228.
- Farzan, A. N., Noack, R., Beachum, L., Taylor, A., Iati, M., Bellware, K., . . . Kornfield, M. (2020, 9 21). *CDC reverses statement on airborne transmission of coronavirus, says draft accidentally published*. Retrieved from Washington Post: https://www.washingtonpost.com/nation/2020/09/21/coronavirus-covid-live-updates-us/
- Glenski, M., Weninger, T., & Volkova, S. (2018, 11 21). Propagation From Deceptive News Sources Who Shares, How Much, How Evenly, and How Quickly? *IEEE Xplore*, 1071-1082. doi:10.1109/TCSS.2018.2881071
- Hinshaw, D. (2020, 10 12). *As Covid Cases Surge, More Public-Health Experts Say Lockdowns Aren't the Answer*. Retrieved from Wall Street Journal: https://www.wsj.com/articles/public-health-experts-rethink-lockdowns-as-covid-cases-surge-11602514769

References #2

- Kouzy, R., Jaoude, J. A., Kraitem, A., Alam, M. B., Karam, B., Adib, E., . . . Baddour, K. (2020, 3 13). Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus*. doi:10.7759/cureus.7255
- Ontario Government. (2020, 10 1). *Official Government of Ontario Twitter*. Retrieved from Twitter: https://twitter.com/ongov
- Oren, E., Martinez, L., Hensley, R. E., Jain, P., Ahmed, T., Purnajo, I., . . . Tsou, M.-H. (2020, 7 29). Twitter Communication During an Outbreak of Hepatitis A in San Diego, 2016–2018. *American Journal of Public Health*. doi:10.2105/AJPH.2020.305900
- Pereira-Kohatsu, J. C., Quijano-Sánchez, L., Liberatore, F., & Camacho-Collados, M. (2019, 11 19). Detecting and Monitoring Hate Speech in Twitter. *sensors, 19*(21), 4654. doi:10.3390/s19214654
- Ravi, K. (2015, 6). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based System, 89*, 14-46. doi:10.1016/j.knosys.2015.06.015
- Sewalk, K. C., Tuli, G., Hswen, Y., Brownstein, J. S., & Hawkins, J. B. (2018, 10). Using Twitter to Examine Web-Based Patient Experience Sentiments in the United States: Longitudinal Study. *Journal of Medical Internet Research, 20*(10). doi:10.2196/10043
- Shao, C., Hui, P.-M., Wang, L., Jiang, X., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2018, 4). Anatomy of an online misinformation network. *PLoS One, 13*(4). doi:10.1371/journal.pone.0196087
- Sicilia, R., Giudice, S. L., Pei, Y., Pechenizkiy, M., & Soda, P. (2018, 11 15). Twitter rumour detection in the health domain. *Expert Systems With Applications, 110*, 33-40. doi:10.1016/j.eswa.2018.05.019
- Vicarioa, M. D., Bessib, A., Zolloa, F., Petronic, F., Scalaa, A., Caldarellia, G., . . . Quattrociocchia, W. (2016, 1 19). The spreading of misinformation online. *Proceedings of the National Academy of Sciences - PNAS, 113*(3), 554-9. doi:10.1073/pnas.1517441113
- Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, a. R. (2018). Detection and Resolution of Rumours in Social Media: A Survey. *ACM Computing Surveys, 51*(2). doi:10.1145/3161603
- Alvas. (2016). How is the Vader 'compound' polarity score calculated in Python NLTK?. Retrieved from: https://stackoverflow.com/questions/40325980/how-is-the-vader-compound-polarity-score-calculated-in-python-nltk