Foster Mosden
MKT 680 - Marketing Analytics
Prof. Ibragimov
Homework Assignment 3 - April 16th, 2024

Setup

To start this assignment, I loaded my data into Python and created dummy variables for the columns ‘platform’, ‘weekday’, and ‘time’, as the questions sought knowledge on the influence of specific attributes from each of those columns. As well, I extracted the numerical values from ‘website_id’, as the column prefaced all of the values with “siteid_”, which made the data otherwise unusable (model seeks integer values). I then created a train/test split, with the test size being 0.2. Finally, I built the LogisticRegression model in Python using the liblinear solver (other solvers failed to converge). From this model I got the following results on the test set:

- Accuracy: 0.8929158624633641
- Precision: 0.8929158624633641
- Recall: 1.0
- F1 Score: 0.943429003021148
- MAE: 0.10708413753663593

Satisfied with the results of this basic yet effective model, I moved on and extracted the coefficients for each feature of the dataset. These coefficients are what I used to answer the questions. You can find my full list of coefficients at the bottom of this document.

Questions

1. Based on the coefficients I observed, consumers are considerably more likely to complete reading a paragraph and thus have increased reading depth on Desktop devices compared to all other platforms. The coefficient for the Desktop dummy variable, ~0.49, is the highest among platform coefficients, which indicates the strongest positive association between Desktop reading and increased reading depth. Mobile came in second place with a coefficient of ~0.31, far behind Desktop. The remaining devices had considerably lower coefficients. It’s also worth noting that of all features in the dataset, Desktop reading had the highest coefficient by far.
2. Again using the coefficients, we observe that the time where users are most likely to finish the paragraph is in the early morning (~0.34). However, the coefficients for all times are fairly close, with hour_evening (~.33) being a very close second in particular. It is worth noting that all of the time based attributes had the 2nd, 3rd, and 4th highest coefficients out of all features in the dataset. As for days of the week, it seems users are most likely to finish the paragraph on Thursdays (~0.22) and Fridays (~0.21).
3. It’s clear from the model that topic 5 (~0.24) and topic 21 (~0.23) are the most highly associated with increased reading depth. All other topics have a coefficient below 0.11, and some (1, 3, 17, and 22) even have negative coefficients.
4. Controlling for topics, the presence of sadness (~0.07) is associated with increased reading depth, as indicated by the positive coefficient. While fear has a small coefficient (~0.009), it is at least positive -- the presence of posemo and/or anger is associated with decreased reading depth, as indicated by their negative coefficients.

5. To build a recommender system, I can use the information about users' reading behaviors (such as device type, time/weekday preferences, topic preferences, emotional language impact) to recommend articles tailored to their preferences. I could use collaborative filtering techniques to recommend articles similar to those read by users with similar behavior profiles. Additionally, I might use content-based filtering by recommending articles based on their topics and emotional content. Finally, I could incorporate user feedback to continuously improve the recommender system's accuracy and effectiveness.

Device Influence:

Feature	Coefficient
platform_Desktop	0.491900
platform_Mobile	0.311001
platform_Tablet	0.236173
platform_Unknown	0.017181

Emotional Language Influence:

Feature	Coefficient
sadness	0.071639
fear	0.008791
posemo	-0.009019
anger	-0.022011

Time Influence:

Feature	Coefficient
time_hour_early_morning	0.341231
time_hour_evening	0.331692
time_hour_afternoon	0.316290
weekday_4	0.221124
weekday_5	0.209381
weekday_6	0.172903
weekday_1	0.146928
weekday_2	0.110767
weekday_3	0.101836
weekday_7	0.093315
time_hour_morning	0.067043

Topic Influence:

Feature	Coefficient
post25_include_5	0.244229
post25_include_21	0.227263
post25_include_4	0.101898
post25_include_2	0.067327
post25_include_15	0.063507
post25_include_20	0.060253
post25_include_9	0.049336
post25_include_25	0.046972
post25_include_12	0.045019
post25_include_23	0.042929
post25_include_13	0.029902
post25_include_14	0.029486
post25_include_11	0.023918
post25_include_18	0.020938
post25_include_7	0.018679
post25_include_16	0.018649
post25_include_6	0.018299
post25_include_8	0.015802
post25_include_10	0.013337
post25_include_24	0.008778
post25_include_19	0.001457
post25_include_1	-0.005263
post25_include_3	-0.012613
post25_include_17	-0.014642
post25_include_22	-0.059203