**Web Appendix A Text Content Classification**

In this paper, we use the standard bag-of-features framework to classify text content into one of the three sentiment types – positive, negative, or neutral, one of the six of topics – three public topics and three personal topics, and one of two behavioral drivers – self-enhancement and providing useful information. Due to Twitter’s 140-character limit, a Tweet usually only covers one single topic. Therefore, supervised machine learning models are more appropriate in our setting than unsupervised methods such as Latent Dirichlet allocation (LDA) which are often used for documents with various topics.

The complete machine learning process proceeds as follows. First, two research assistants manually labeled 2,000 Tweets for training purposes. Regarding the topics, we first differentiate public information/news from personal experiences/opinions. We then control for the two most common sub-topics of each type based on the result of manual coding on our training sample of 2,000 Tweets, and group the remaining public topics into “Other Public Topics” and the remaining personal topics into “Other Personal Topics.” For example, in our airline Tweets sample, the two most common public topics are “9-11” and “Ticket Sale,” and the two most common personal topics are “Luggage or Delay” and “In-Flight Experience.” In the second step, we define the “feature bag” to be processed by the classifier. This is done by first tokenizing the Tweets into sets of unigram and bigram features (the sequence of features is ignored) and filtering out all stopping words and punctuations (except question marks and exclamation points) because they contain little information about the variables. Then, a measure called Term Frequency, Inverse Document Frequency, abbreviated to tf-idf is calculated for each text feature using the TfidfVectorizer function of *scikit-learn*, a machine learning library for the Python programming language. Tf-idf calculates the relative frequency of a feature for a single Tweet, compared to the inverse presence of the feature throughout an entire corpus of Tweets. Using tf-idf score instead of an indicator of the presence of a feature allows us to capture how important a word is to a Tweet among all Tweets (Ramos, 2003). In the third step, the tf-idf vector and labels are fed into four classification algorithms to generate a model. These four algorithms include: 1) Linear Support Vector Machine, 2) Multinomial Naïve Bayes, 3) Logistic Regression, and 4) Random Forest Classifier. We have evaluated their performance using 10-fold cross-validation. The average accuracy rates for that Airline Tweets dataset are reported in Table A.1. Relevant statistics for the other two datasets are available upon request. The results show that the Linear Support Vector Machine outperforms the other three algorithms consistently. Therefore, we choose Linear SVM as our classification method. Finally, the same feature extractor is used to convert remaining Tweets in our sample to a tf-idf vector of features. These feature sets are then fed into the model which generates the predicted labels.

**Table A-1 Comparison of 10-Fold Cross-Validation Accuracy Rate**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Personal* | *Sentiment* | *Useful Information* | *Good Image* |
| Linear SVM | .891 | .714 | .869 | .927 |
| Multinomial NB | .885 | .640 | .793 | .734 |
| Logistic Regression | .857 | .631 | .835 | .910 |
| Random Forest Classifier | .655 | .466 | .791 | .910 |

**Web Appendix B Propensity Score Matching for Twitter Datasets**

In this section, we present the parameter estimates for the propensity score model for the airline Tweets sample and the SUV Tweets sample. The dependent variable is whether a Tweet contains imagery content. The independent variables include features of text content as well as post and poster level characteristics. We do not perform propensity score matching on image characteristics or image-text relevancy because the resulting matching sample would be too small to yield credible inferences. Controlling for the presence of imagery content is adequate because it is hard to imagine a poster is not strategic when deciding whether to attach a picture or not in the first place but is strategic when deciding which picture to attach. We also do not differentiate directly viewable pictures from linked pictures in the propensity score analysis because it is difficult to predict whether the poster will post the picture directly or share a picture link based on observable post and poster characteristics.

**Table B-1 Estimation Results for Propensity of Imagery Content: Airline Tweets**

|  |  |  |
| --- | --- | --- |
| *Variables* | *Levels* | *Parameter Estimates* |
| Intercept |  | **-.29 (.12)**  |
| Shutterbug |  | **.26 (.10)**  |
| Tweets-with-Image Percentage |  | **4.48 (.14)**  |
| Log of Number of Followers |  | **.05 (.01)**  |
| Verified Account |  | **.44 (.09)**  |
| Sentiment | Positive | **.14 (.05)**  |
|  | Negative | **-.39 (.06)**  |
| Topics | Ticket Sale | **-.74 (.10)**  |
|  | Sept-11 | **-.62 (.10)**  |
|  | Delay and Luggage | **-.10 (.11)**  |
|  | Flight Experience | **-1.10 (.07)**  |
|  | Other Personal | **-.70 (.06)**  |
| Behavioral Drivers | Useful Information | **-.91 (.06)** |
|  | Good Image | .18 (.14)  |
| Brands | JetBlue | **.61 (.06)**  |
|  | Frontier | **-.36 (.16)**  |
|  | Southwest | **.43 (.08)**  |
|  | Virgin | **.79 (.11)**  |
|  | Delta | **-.73 (.11)**  |
|  | United | **-1.16 (.09)**  |
|  | Alaska | **-1.03 (.26)**  |
|  | Spirit | **-.42 (.21)**  |
| At (@) |  | **-.08 (.03)**  |
| Hashtag |  | **.25 (.02)**  |
| Emoji |  | **.06 (.02)**  |
| Number of Words |  | **-.08 (.00)**  |
| Linguistic Content Category |  | **🗸** |
| Time-of-day fixed effects |  | **🗸** |
| Weekend fixed effects |  | **🗸** |

**Table B-2 Estimation Results for Propensity of Imagery Content: SUV Tweets**

|  |  |  |
| --- | --- | --- |
| *Variables* | *Levels* | *Parameter Estimates* |
| Intercept |  | **-3.51 (.19)**  |
| Shutterbug |  | **1.14 (.15)**  |
| Tweets-with-Image Percentage |  | **4.35 (.11)**  |
| Log of Number of Followers |  | **.14 (.01)**  |
| Verified Account |  | **.80 (.16)**  |
| Dealer Account |  | **.19 (.09)**  |
| Other Auto-Related Account |  | **.23 (.08)**  |
| Sentiment | Positive | **.61 (.07)**  |
|  | Negative | -.07 (.18)  |
| Topics | Car Sale | **-.21 (.10)**  |
|  | Recall | -.49 (.32)  |
|  | New Purchase | -.02 (.15)  |
|  | Driving Experience | **.71 (.14)**  |
|  | Other Personal | **.84 (.10)**  |
| Behavioral Drivers | Useful Information | **.48 (.09)**  |
|  | Good Image | **.72 (.13)**  |
| Brands | Chevy | **-.46 (.16)**  |
|  | Ford | -.08 (.12)  |
|  | Honda | .05 (.13)  |
|  | Jeep | -.14 (.11)  |
|  | Nissan | **-.47 (.14)**  |
|  | Subaru | **-1.33 (.13)**  |
| At (@) |  | -.03 (.05)  |
| Hashtag |  | **.25 (.02)**  |
| Emoji |  | -.07 (.07)  |
| Number of Words |  | **-.07 (.01)**  |
| Linguistic Content Category |  | **🗸** |
| Time-of-day fixed effects |  | **🗸** |
| Weekend fixed effects |  | **🗸** |

**Web Appendix C Summary Statistics for Airline Instagram Sample**

**Table C-1 Summary Statistics: Airline Instagram (Discrete Variables)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Levels* | *Count* | *Percent* |
| Relevant | Yes | 1,989 | 97.31% |
|  | No | 55 | 2.69% |
| Human Image | Face Presence | 268 | 13.11% |
|  | Happy Face | 157 | 7.68% |
|  | Other or No Expression | 111 | 5.43% |
| Source | Screenshot | 162 | 7.95% |
|  | Amateur Photo | 1,527 | 74.71% |
|  | Professional Photo | 355 | 17.34% |
| Quality | High | 891 | 43.59% |
|  | Medium or Low | 1,153 | 56.41% |
| Sentiment | Positive  | 611 | 29.89% |
|  | Negative  | 312 | 15.26% |
|  | Neutral | 1,121 | 54.85% |
| Behavioral Driver | Useful Information | 75 | 3.67% |
|  | Good Image | 59 | 2.89 |
| Personal Topic | Yes | 1,749 | 85.57% |
|  | No | 295 | 14.43% |
| Brands | AA | 466 | 22.80% |
|  | JetBlue | 173 | 8.46% |
|  | Frontier | 24 | 1.17% |
|  | Southwest | 307 | 15.02% |
|  | Virgin | 12 | .59% |
|  | Delta | 341 | 16.68% |
|  | United | 303 | 14.82% |
|  | Alaska | 133 | 6.51% |
|  | Spirit | 42 | 2.05% |

**Table C-2 Summary Statistics: Airline Instagram (Continuous Variables)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Mean* | *S.D.* | *Min* | *Max* |
| Log Number of Likes | 4.02 | 1.05 | 2.40 | 9.37 |
| Top 3 Color % | 45.8% | 20.6% | 1.3% | 99.5% |
| Number of Mentions | .56  | 1.84  | 0 | 27 |
| Number of Hashtags | 12.62  | 9.57  | 0 | 70 |
| Number of Emojis | .28  | 1.11  | 0 | 16 |
| Number of Words | 19.57  | 31.42  | 0 | 379 |
| Log Number of Followers | 6.57  | 1.43  | 1.95  | 13.03  |

**Web Appendix D Robustness Check on Poster Level Effect**

To alleviate the concern that the effects of imagery content might be driven by some heavy social media users, we have taken a closer look at the poster level effects. In the matched sample for the airline Tweets, we have 6,830 posts from 5,205 posters/accounts. There are 619 (11.8%) posters who had published at least two posts in our airline twitter sample. To include 619 poster level fixed effect for these posters is not doable given the non-linear nature of our main model. Instead, we have run a robustness check by including poster-specific intercepts for those accounts that have posted more than five times in our matched sample. We have 78 such posters, and therefore we included 156 fixed effects (78 for liking and 78 for retweeting) in our main model to assess the robustness of our main findings of poster level effects.

The estimation results are presented in Table D-1. By comparing Table D-1 to Table 4 in the paper, we can see that the parameter estimates are qualitatively unchanged for all main variables of interest after we include the poster level fixed effect. Therefore, we can conclude that our main findings are not due to posters who are heavy users of social media and are robust to the inclusion of poster level effects.

**Table D-1 Estimation Results for User Engagement:**

**Airline Tweets with Poster Level Dummies**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | *Attention* | *Retweets* | *Likes* |
|  | Intercept | 1.415\*\*\*(.115) | -4.150\*\*\*(.072) | -2.932\*\*\*(.064) |
| **Imagery Content** |  |  |  |  |
| Mere Presence | Direct Image | 1.055\*\*\*(.039) |  1.956\*\*\*(.049) |  .977\*\*\*(.047) |
|  | Linked Image | 3.232\*\*\*(.198) | -2.428\*\*\*(.070) | -1.953\*\*\*(.066) |
| Imagery Characteristics |  |  |  |  |
| *Colorfulness* | Top 3 Color | **-----** | -1.019\*\*\*(.043) | .022(.042) |
| *Facial Features* | Human Face | **-----** |  .590\*\*\*(.043) |  .362\*\*\*(.040) |
|  | Happy Face | **-----** | -.669\*\*\*(.050) | .156\*\*\*(.051) |
| *Picture Source* | Screenshot | **-----** | -.261\*\*\*(.038) | -.010(.040) |
|  | Amateur Photo | **-----** | -.347\*\*\*(.044) | -.195\*\*\*(.041) |
| *Picture Quality* | High Quality | **-----** |  .184\*\*\*(.067) | .034 (.061) |
| Image-Text Fit | Relevancy | **-----** | -.118 (.067) |  .187\*\*\*(.044) |
| **Text Content** |  |  |  |  |
| Sentiment | Positive Sentiment | **-----** |  .417\*\*\*(.041) |  .182\*\*\*(.040) |
|  | Negative Sentiment | **-----** |  .354\*\*\*(.047) | -.217\*\*\*(.044) |
| Topics | Ticket Sale | **-----** |  -.467\*\*(.058) | -.165\*\*\*(.076) |
|  | 9-11 | **-----** |  .654\*\*\*(.078) | -.118(.077) |
|  | Delay and Luggage | **-----** | -2.020\*\*\*(.106) | -.330\*\*\*(.076) |
|  | Flight Experience | **-----** | -1.020\*\*\*(.062) | -.089(.058) |
|  | Other Personal Topics | **-----** | -.948\*\*\*(.044) | -.222\*\*\*(.043) |
| Drivers | Useful Information | **-----** | -.370\*\*\*(.051) | -1.069\*\*\*(.048) |
|  | Good Image | **-----** |  1.015\*\*\*(.035) |  .804\*\*\*(.037) |
| Other Text Features | Number of Mentions (@) |  4.652\*\*\*(.161) |  .159\*\*\*(.031) |  .156\*\*\*(.028) |
|  | Number of Hashtags | -.279\*\*\*(.073) |  .134\*\*\*(.025) |  .077\*\*\*(.023) |
|  | Number of Emojis | 3.049\*\*\*(.052) | -.157\*\*\*(.036) | .080\*\*\*(.026) |
|  | Number of Words | .041(.023) | .026\*\*\*(.005) |  .025\*\*\*(.005) |
| Brand Mentioned | JetBlue | **-----** | -.465\*\*\*(.054) | -.429\*\*\*(.049) |
|  | Frontier | **-----** | -.623\*\*\*(.077) | .060(.071) |
|  | Southwest | **-----** | -.819\*\*\*(.054) | -.235\*\*\*(.057) |
|  | Virgin | **-----** | -.493\*\*\*(.034) | -.107\*\*\*(.035) |
|  | Delta | **-----** | -.437\*\*\*(.048) | -.597\*\*\*(.047) |
|  | United | **-----** | -.426\*\*\*(.038) | -1.062\*\*\*(.039) |
|  | Alaska | **-----** | -.471\*\*\*(.010) | -.683\*\*\*(.013) |
|  | Spirit | **-----** | -.579\*\*\*(.047) | -.773\*\*\*(.057) |
| **Account Characteristics and Other Controls** |  |  |  |
| Account Characteristics | Log Number of Followers | .052(.039) |  .442\*\*\*(.010) | .378\*\*\*(.009) |
| Verified Account | 5.489\*\*\*(.147) | .832\*\*\*(.062) | 1.159\*\*\*(.057) |
| Linguistic Features | Linguistic Content Category | **-----** | **🗸** | **🗸** |
| Poster Fixed Effects |  | **-----** | **🗸** | **🗸** |
| Time Fixed Effects | Time-of-day Fixed Effects | **🗸** | **-----** | **-----** |
|  | Weekend Fixed Effects | **🗸** | **-----** | **-----** |
| **Correlation Between Sharing and Liking** |
|  | Alpha | 2.499\*\*\*(.055) |  |  |

Note: Standard errors in parentheses; \*\*\**p* < .01; \*\**p* < .05;