Table 1: Descriptive statistics of the organisation groups.

	Telecom	Sci	Fin	Gov
Tweets: engaged (% of all tweets)	2827 (61.87)	4400 (**86.85**)	1115 (83.77)	1685 (75.06)
Retweets (%)	1035 (36.61)	3813 (**86.66**)	918 (82.33)	1302 (77.27)
Replies (%)	2351 (**83.16**)	864 (19.64)	459 (41.17)	533 (31.63)
Likes (%)	1476 (52.21)	4137 (**94.02**)	1007 (90.31)	1357 (80.53)
Tweets: not engaged (%)	1742 (38.13)	666 (13.15)	216 (16.23)	560 (24.94)

Table 2: ANOVA results. : indicates interaction. *** denotes $\alpha = 0.001$ and ** denotes $\alpha < 0.01$.

Predictor	F value	Probability	Significance
org. type	50.9741	< 2.2e-16	***
hashtag:org. type	87.4272	< 2.2e-16	***
url:org. type	3.7362	0.0048329	**
photo:org. type	34.2711	< 2.2e-16	***
video:org. type	17.0346	6.095e-14	***
exclamation point:org. type	5.0327	0.0004736	***
dominance:org. type	3.9185	0.0035019	**
time:org. type	2.6461	0.0015244	**
animation	34.5361	4.286e-09	***
arousal	7.3233	0.0068155	**

interaction of these variables with a special predictor variable, "Organisation Type", as outlined above. The Twitter content-oriented predictors are:

- **Twitter metadata** (hashtag, mention, URL, photo, video, animation, timestamp): These are predictors about the tweet itself but not the tweet text. Tweet's timestamp is discretised into four labels: morning, daytime, evening, and night.
- **Stylistic** (contraction, abbreviation, slang, exclamation point, question mark, capitalised word, lowercase word, first pronoun, second pronoun, third pronoun): we incorporate stylistic predictors from tweet text motivated by prior work [6]. We use a dictionary [7] to identify slang words, abbreviations, contractions or emoticons, and we employ the CMU Twitter POS tagger [5] to acquire personal pronouns from the text.
- **Sentiment-Emotion** (sentiment-phrases, valence, arousal, dominance): Predictors are generated from two emotion-sentiment lexicons: "ANEW" [2] and "PERMA" [8]. The lexicons provide scores per word which are summed and normalised.

4. REGRESSION ANALYSIS

We quantitatively evaluate the effects of the factors that significantly impact on tweet engagement by performing statistical hypothesis testing using a linear regression. We first use a forward and backward stepwise model-search algorithm with the Akaike Information Criterion (Stepwise-AIC) to eliminate non-significant interaction predictor variables between organisation type and content-oriented variables. For the remaining predictors, we obtain a linear regression model and then perform an Analysis of Variance (ANOVA).

The results of the ANOVA are presented in Table 2. We observe that there is a main effect with the organisation type, which is a significant predictor of engagement levels. The predictors showing significant interactions with organisation type are shown in the middle row. Finally, we also observe two main effects for which there are no interactions with organisation type: the animation and arousal predictor variables.

5. DISCUSSION OF RESULTS

The regression analysis suggests that, for this data set, one generally needs to consider the organisation type before selecting strategies for deciding upon the contents of a Twitter post. The exceptions are the use of animation, and the use of words that scores highly for the emotional arousal, both of which seem to be strongly associated with engagements regardless of organisation type. The latter is consistent with other research [4]. However, we find that the use of words with emotional Dominance (weakness vs. strength of emotions) is dependent on organisation type.

Content creation strategies should thus take into account the organisation type. We illustrate this with one example but leave a full analysis to future work. Consider the case of video content. For Finance organisations, including such content has a coefficient of 1.418 ($p<0.001$), whereas for Government departments, the coefficient is -0.551 ($p<0.05$), indicating that while video content is more associated with engagements in the Finance domain, including video content may not help engagement on Government Twitter posts.

6. CONCLUSIONS

Organisations are increasingly interested in understanding how to create content on social media to maximise engagement by target audiences. In this work, we examined whether the type of organisation had an impact on how to write tweets to attract engagement. Our results show this is the case: tweet characteristics for predicting engagement vary depending on the organisation types. That is, content creation strategies are not universal. This is important as it shows that the strategies one uses to create Twitter content should depend on the organisation type.

7. REFERENCES

[1] C. Bauckhage, K. Kersting, and F. Hadiji. Mathematical Models of Fads Explain the Temporal Dynamics of Internet Memes. *ICWSM*, 2013.

[2] M. M. Bradley and P. J. Lang. Affective norms for english words (ANEW): Instruction manual and affective ratings. Technical report, 1999.

[3] Y. Hu, S. Farnham, and K. Talamadupula. Predicting user engagement on twitter with real-world events. *ICWSM*, 2015.

[4] N. Naveed, T. Gottron, J. Kunegis, and A. Che Alhadi. Bad news travel fast: A content-based analysis of interestingness on twitter. *WebSci*, 2011.

[5] O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, N. Schneider, and N. A. Smith. Improved part-of-speech tagging for online conversational text with word clusters. *NAACL-HLT*, pages 380–390, 2013.

[6] C. Paris, P. Thomas, and S. Wan. Differences in language and style between two social media communities. *ICWSM*, 2012.

[7] J.-P. Posadas-Durán, I. Markov, H. Gómez-Adorno, G. Sidorov, I. Batyrshin, A. Gelbukh, and O. Pichardo-Lagunas. Syntactic n-grams as features for the author profiling task. *CLEF*, 2015.

[8] H. A. Schwartz, M. Sap, M. L. Kern, J. C. Eichstaedt, A. Kapelner, M. Agrawal, E. Blanco, L. Dziurzynski, G. Park, D. Stillwell, et al. Predicting Individual Well-Being Through the Language of Social Media. *PSB*, 2016.