Success Prediction of Instagram Posts

Authors…

Summary

[Abstract 3](#_Toc94892575)

[Introduction 3](#_Toc94892576)

[Text pre-processing 3](#_Toc94892577)

[References 5](#_Toc94892578)

# Abstract

In this work we aim to develop a predictor based on NLP understanding for social media content optimisation. We would like to prove how the right hashtags, tags and caption can drive the success of a post both from a sentiment analysis comments perspective and number of likes and comments. This is a use case that could be interesting in a real-life scenario, for influencers that would like to maximise their impact on the platform. We plan on using the same model and train it on a different subset of the data, which will be chosen accordingly by a clustering model to differentiate Instagram profile of different nature. The proposed approach exploits big data technologies for scalability and efficiency, and it is general enough to be applied to other social media as well.

# Introduction

Instagram is a photo-sharing platform that was launched in 2010. It has gradually gained a leading role among photo-sharing platforms, introducing several innovative features over times, including filters, stories, and an internal messaging system. These features have attracted not only ordinary users and photography enthusiasts, but also companies, organizations, and global brands, thanks to the possibility that Instagram has offered to explore new business models and marketing strategies. In our work we use a comprehensive dataset of Instagram posts from many different successful profiles and the following steps will be done:

* Text pre-processing
* Data exploration
* …
* …
* …
* …
* Error Analysis
* Final Conclusion

However, we can scan our work in two macro steps: The Feature Engineering Stage and Supervised Learning Step [Figure 1]. The first includes the data pre-processing, thus we enriched the dataset with some derived information, as well as removing data whose contribution was negligible or not interesting for the class prediction purposes. Instead, the Supervised Learning Step includes the classification model training and the success prediction of new posts.

*Supervised learning stage*

*Feature engineering stage*

Figure 1 Workflow

# Data pre-processing

To prepare the text data for the model building we perform text preprocessing. It is the very first step of NLP projects. First we converted the emojis from Unicode format to simple text, thanks to the *emojis.unicode\_codes* library.

…

We extract the hashtags from the caption string for each post and we create another column *‘hashtags’* which is useful to predict the likes. Afterwards we apply the typical pre-processing tasks as the lower casing, removal of punctuation, removal of frequent and rare words, stemming, lemmatization. All the functions that develop these changes are implemented in the python file *‘preprocessing\_functions.py*’.

# References

<https://towardsdatascience.com/predicting-the-popularity-of-instagram-posts-deeb7dc27a8f>