Kiki Kills: Identifying Dangerous Challenge Videos from Social Media

Nupur Baghel
MIDAS Lab, NSIT-Delhi
nupurb.co@nsit.net.in

Yaman Kumar
Adobe Systems
ykumar@adobe.com

Paavini Nanda
MIDAS Lab, NSIT-Delhi
paavinin.co@nsit.net.in

Rajiv Ratn Shah
IIIT, Delhi
rajivratn@iiitd.ac.in

Debanjan Mahata
University of Arkansas at Little Rock
dxmahata@ualr.edu

Roger Zimmermann
NUS, Singapore
rogerz@comp.nus.edu.sg

Abstract

There has been upsurge in the number of people participating in challenges made popular through social media channels. One of the examples of such a challenge is the *Kiki Challenge*, in which people step out of their moving cars and dance to the tunes of the song, *"Kiki, Do you love me?"*. Such an action makes the people taking the challenge prone to accidents and can also create nuisance for the others traveling on the road. In this work, we introduce the prevalence of such challenges in social media and show how the machine learning community can aid in preventing dangerous situations triggered by them by developing models that can distinguish between dangerous and non-dangerous challenge videos. Towards this objective, we release a new dataset namely *MIDAS-KIKI dataset*, consisting of manually annotated dangerous and non-dangerous Kiki challenge videos. Further, we train a deep learning model to identify dangerous and non-dangerous videos, and report our results.

Introduction

With the penetration of internet in masses, various types of uploaded content in social media have set new trends. Previous trends, like advertising (Miller and Lammas 2010), crisis reportage (Sakaki, Toriumi, and Matsuo 2011), dangerous selfies (*killfies*) (Lamba et al. 2016) have been thoroughly studied. There has also been some work on challenges like *Blue Whale* (Mukhra et al. 2017). However, a challenge of the magnitude and gravity such as *Kiki Challenge* has not been studied previously. This is partly due to it being a very recent phenomenon, and the other reason being the unavailability of public datasets for such video-based challenges.

The *Kiki challenge*, involves people dancing to the lyrics of the song, *"Kiki, Do you love me?"* and filming themselves, while a car is moving on the road. This poses a serious risk for the performers besides the people who might be using the road at the same time. Several cases of traffic blockages, injuries, accidents, crashes and even deaths have been reported relating to this challenge. Abu Dhabi, Spain and Egypt governments have started prosecution and arrest with heavy fines against the offenders[2].

Insensitive promotion of this challenge by celebrities have been encouraging this hysteria. For instance, Will Smith per-

formed Kiki dance on top of Budapest bridge[1], thus effectively encouraging his 77 million followers to do the same. There have been arrests[2], some known and reported deaths[3], multiple cases of acute injuries[4] due to this worrying trend.

Contributions - In this paper, we characterize this problem, collect the relevant videos from social networking websites, release the dataset for public use and propose and build a model which can differentiate dangerous Kiki videos from the non-dangerous ones. Furthermore, we believe that our contributions would provide a tool for automatic analysis of videos such as those in Kiki challenge to reduce the number of deaths and injuries. To promote research in this domain, we also release the dataset (and metadata) that we collect for this purpose at http://midas.iiitd.edu.in.

Methodology

Dataset - We collected data from Twitter, using its publicly available APIs. We scraped the tweets and selected those containing at least one of the hashtags: *#kiki, #keke, #kikichallenge, #kekechallenge, #inmyfeelings, #kikidoyouloveme, #kekedoyouloveme, #inmyfeelingschallenge, #shiggychallenge and #drake*. In total, we found 703 hashtags which were prevalent in tweets related to Kiki challenge. In this process, we collected more than 25k tweets, over a time period of three months from 20th June to 10th September, 2018. This period was chosen since dancing to the tunes of the song became a sensation only after the comedian, *Shiggy* did the same in the later part of June.

After filtering tweets containing videos we finally obtained 2000 kiki challenge videos. Two annotators categorized these videos into *dangerous* and *non-dangerous*. Out of the total videos, 220 videos were tagged dangerous and the rest were non-dangerous. For categorizing a video as dangerous, the annotators were asked to answer a single question, *"After watching this video, does police and/or ambulance need to be informed?"* There was a random assignment of videos to the two annotators. The inter-annotator agreement,

---

[1] https://www.bbc.com/news/newsbeat-44817582
[2] https://bit.ly/2NK7DLU, https://bit.ly/2NLf4CI
[3] https://bit.ly/2MwR1T7
[4] https://wapo.st/2Qx9uSR

Figure 1: An example for non-dangerous Kiki performance

Figure 2: An example for dangerous Kiki performance

as calculated using Cohen's Kappa was found out to be *0.94*, which suggests a good agreement between the annotators. The risk perceived by annotators in most cases was from a person being hit by another car, falling on the road or getting hit by a pole while dancing, and in some cases, of overt acts of aggression by the performers, such as burning the car. A sample of dangerous and non-dangerous video are shown in Figures 1 and 2.

Model - We trained a deep learning model using transfer learning on the annotated dataset. The architecture of the model is shown in Figure 3. The input video frames are fed to a VGG-16 (Simonyan and Zisserman 2014) model wrapped in a TimeDistributed wrapper in Keras in conjugation with a flattened layer followed by several dense and dropout layers (with dropout as 0.5). The last layer decides whether the video fed to the system was dangerous or not. In order to train the model, the dataset was splitted into three parts: 60% for training, 20% for validation and 20% for testing using stratified sampling. The validation set was used for deciding the architecture and parameters of the final model. After training we tested our model on the test data, obtaining and accuracy of 87%. The performance metrics of the model is reported in Table 1.

Table 1: Performance metrics for our model.

Accuracy	Precision	Recall	F1-Score
0.87	0.96	0.90	0.93

Conclusion

In this paper, we characterized the recent trend of taking widely popular challenges on social media platforms that can create safety hazards and life-threatening situations. We took *Kiki Challenge* as our use case, and implemented a deep learning model on videos that differentiates between dangerous and non-dangerous Kiki Challenge videos collected from Twitter. Our model shows high accuracy in identifying
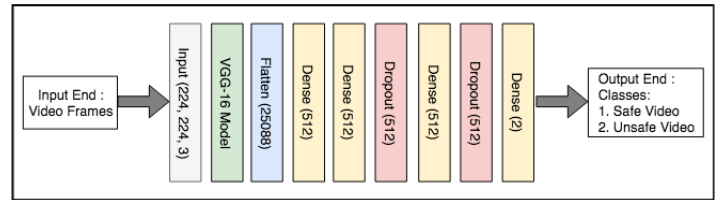

Figure 3: Model architecture for classification of Kiki challenge videos.

the two classes. In this process, we compiled a manually annotated dataset of dangerous and non-dangerous videos, and make it publicly available. We believe that our work would encourage researchers to use machine learning for identifying potentially dangerous challenge videos, and promote public safety. As part of our future work, we would like to investigate the problem domain in greater detail by taking into account the data imbalance and considering other metadata associated with the videos, while training our models.

References

Lamba, H.; Bharadhwaj, V.; Vachher, M.; Agarwal, D.; Arora, M.; and Kumaraguru, P. 2016. Me, myself and my killfie: Characterizing and preventing selfie deaths. Technical report.

Miller, R., and Lammas, N. 2010. Social media and its implications for viral marketing. *Asia Pacific Public Relations Journal* 11(1):1–9.

Mukhra, R.; Baryah, N.; Krishan, K.; and Kanchan, T. 2017. blue whale challenge: A game or crime? *Science and engineering ethics* 1–7.

Sakaki, T.; Toriumi, F.; and Matsuo, Y. 2011. Tweet trend analysis in an emergency situation. In *Proceedings of the Special Workshop on Internet and Disasters*, 3. ACM.

Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.