DANKMEMES @ EVALITA 2020:
The Memeing of Life:
Memes, Multimodality and Politics

Martina Miliani[1,2] and **Giulia Giorgi**[3] and **Ilir Rama**[3]
and **Guido Anselmi**[3] and **Gianluca E. Lebani**[4]

[1] University for Foreigners of Siena
[2] CoLing Lab, Department of Philology, Literature, and Linguistics, University of Pisa
[3] Department of Social and Political Sciences, University of Milan
[4] Department of Linguistics and Comparative Cultural Studies, Ca' Foscari University of Venice

```
martina.miliani@fileli.unipi.it, giulia.giorgi@unito.it
ilir.rama@unimi.it, guido.anselmi@unimi.it,
gianluca.lebani@unive.it
```

Abstract

DANKMEMES is a shared task proposed for the 2020 EVALITA campaign, focusing on the automatic classification of Internet memes. Providing a corpus of 2.361 memes on the 2019 Italian Government Crisis, DANKMEMES features three tasks: A) Meme Detection, B) Hate Speech Identification, and C) Event Clustering. Overall, 5 groups took part in the first task, 2 in the second and 1 in the third. The best system was proposed by the UniTor group and achieved a F_1 score of 0.8501 for task A, 0.8235 for task B and 0.2657 for task C. In this report, we describe how the task was set up, we report the system results and we discuss them.

1 Introduction

Internet memes are understood as "pieces of culture, typically jokes, which gain influence through online transmission" (Davison, 2012). Specifically, a meme is a multimodal artefact manipulated by users, who merges intertextual elements to convey an ironic message. Featuring a visual format that includes images, texts or a combination of them, memes combine references to current events or relatable situations and pop-cultural references to music, comics and movies (Ross and Rivers, 2017).

The pervasiveness of meme production and circulation across different platforms increases the necessity to handle massive quantities of visual data (Tanaka et al., 2014) by leveraging on automated approaches. Efforts in this direction focused on the generation of memes (Peirson V and Tolunay, 2018; Gonçalo Oliveira et al., 2016) and on automated sentiment analysis (French, 2017), while stressing the need for a multimodal approach able to contextually consider both visual and textual information (Sharma et al., 2020; Smitha et al., 2018).

As manual labelling becomes unfeasible on a large scale, scholars require tools able to classify the huge amount of memetic content continuously produced on the web. The main goal of our shared task is to evaluate a range of technologies that can be used to automatize the process of meme recognition and sorting with an acceptable degree of reliability.

2 Task Description

The DANKMEMES task, presented at the 2020 EVALITA campaign (Basile et al., 2020), encompasses three subtasks, aimed at: detecting memes (Task A), detecting the hate speech in memes (Task B) and clustering memes according to events (Task C). Participants could decide to take part in one or more of these tasks, with the only recommendation that Task 1 functions as the compulsory preliminary step for the other two tasks.

Task A: Meme Detection. The lack of consensus around what defines a meme (Shifman, 2013) led to different definitions, focusing on circulation (Davison, 2012; Dawkins, 2016), formal features (Milner, 2016), or content (Gal et al., 2016; Kno-

bel and Lankshear, 2007). For this dataset, manual coding focused both on formal aspects (such as layout, multimodality and manipulation) as well as content, e.g. ironic intent (Giorgi and Rama, 2019); the exponential increase in visual production, however, warrants an automated approach, which might be able to further tap into stable and generalizable aspects of memes, considering form, content and circulation. Given the dataset minus the variable strictly related to memetic status, participants must provide a binary classification, distinguishing memes (1) from non memes (0).

Task B: Hate Speech Identification. Hate speech became a relevant issue for social media platforms. Even though the automatic classification of posts may lead to censorship of non-offensive content (Gillespie, 2018), the use of machine learning techniques became more and more crucial, since manual filtering is a very time consuming task for the annotators (Zampieri et al., 2019b). Recent studies have also shown that multimodal analysis is fundamental in such a task (Sabat et al., 2019). In this direction, SemEval 2020 proposed the "Memotion Analysis" among its tasks, to classify sarcastic, humorous, and offensive meme (Sharma et al., 2020). This kind of analysis assumes a specific relevance when applied to political content. Memes about political topics are a powerful tool of political criticism (Plevriti, 2014). For these reasons, the proposed task aims at detecting memes with offensive content. Following Zampieri (2019a) definition, an offensive meme contains any form of profanity or a targeted offense, veiled or direct, such as insults, threats, profane language or swear words. Thus, the second task consists in a binary classification, where systems have to predict whether a meme is offensive (1) or not (0).

Task C: Event Clustering. Social media react to the real world, by commenting in real-time to mediatised events in a way that disrupts traditional usage patterns (Al Nashmi, 2018). The ability to understand which events are represented and how, then, becomes relevant in the context of an hyper-productive Internet.
The goal of the third subtask is to cluster a set of memes that may be or may be not related to the 2019 Italian government crisis into five event categories (see Table 1).

Participants' goal is to apply supervised tech-

Label	Description
0	Residual category
1	Beginning of the government crisis
2	Conte's speech and beginning of consultations
3	Conte is called to form a new government
4	5SM holds a vote on the platform Rousseau

Table 1: Categories for Task C: Event Clustering.

niques to cluster the memes, so that memes pinpointing to the same events are classified in the same cluster.

3 Dataset

3.1 Composition of the dataset

The DANKMEMES dataset is comprised of 2,361 images (for each subtask a specific dataset was provided), automatically extracted from Instagram through a Python script aimed at the hashtag related to the Italian government crisis ("#crisidigoverno"). The corpus includes 367 offensive political memes unrelated to the government crisis, and aimed at augmenting and balancing the dataset for task 2.

3.2 Annotation of the dataset

For each image of the dataset we provide both the name of the .jpg image file, the date of publication and the engagement, i.e. the number of comments and likes of the post. The dataset also includes image embeddings. The vector representations are computed employing ResNet (He et al., 2016), a state-of-the-art model for image recognition based on Deep Residual Learning. Providing such image representations allows the participants to approach these multimodal tasks focusing primarily on its NLP aspects (Kiela and Bottou, 2014). The annotation process involved two Italian native speakers, who study memes at an academic level, and focused on detecting and labelling 7 relevant categories:

- **Macro status**: refers to meme layouts and their relation to diffused, conventionalised formats called macros. The category has 0 and 1 as labels, where the value 1 represents well-known memetic frames, characters and layouts (e.g. Pepe the Frog). The identification of macros relied both on external sources

(e.g. the website "Know Your Meme") and the annotators' literacy on memes.

- **Picture manipulation**: entails the degree of visual modification of the images. Non-manipulated or low impact changes are labeled 0 (e.g. the addition of a text or a logo). Heavily manipulated, impactful changes (e.g. images edited to include political actors) are labeled 1.

- **Visual actors**: the political actors (i.e. politicians, parties' logos) portrayed visually, regardless whether edited into the picture or portrayed in the original image.

- **Text**: the textual content of the image has been extracted through optical character recognition (OCR) using Google's Tesseract-OCR Engine, and further manually corrected.

- **Meme**: binary feature, where 0 represents non meme images and 1 meme images. This is the target label for Task A.

- **Hate Speech**: binary feature only for memes. It differentiates memes with offensive language (1) from non offensive memes (0). This is the target label for Task B.

- **Event**: it is a feature only for meme images, categorizing them according to 4 events (described in 4), plus a residual category labeled as 0. This is the target label for Task C.

The final inter-annotator agreement (IAA) has been calculated by two of the authors on a subset of the dataset through Krippendorff's alpha (Krippendorff, 2018). Four features have been considered: Macro status ($\alpha = 0.755$), Picture manipulation ($\alpha = 0.930$), Hate Speech ($\alpha = 0.741$) and Meme ($\alpha = 0.884$). Other features were either objective (i.e. Visual and textual actors) or inferred from external data (i.e. events).

Participants were allowed to use external resources, lexicons or independently annotated data. Given that, although we provided ResNet image embeddings, participants could make use of any other image representations.

3.3 Training and Test Data

The initial dataset was split into three datasets, one for each task, structured as follows:

Figure 1: Two examples from the dataset for Meme Detection: the image at the top is a meme, whereas the image at the bottom is not a meme.

Dataset for Meme Detection (Task A). The whole dataset counts 2,000 images, half memes and half not (see Figure 1 for an example). We split the dataset into training and test sets, in a proportion of 80-20% of items. Table 2 represents the format of the training dataset. The test dataset has been provided without gold labels, i.e. without the "Meme" attribute.

Dataset for Hate Speech Identification (Task B). The whole dataset counts 1,000 memes (see Figure 2 for an example). We split the dataset into training and test sets, in a proportion of 80-20% of items. Table 3 represents the format of the training dataset. The test dataset has been provided without the gold label "Hate Speech" for testing purposes.

Dataset for Event Clustering (Task C). The whole dataset counts 1,000 memes (see Figure 3 for an example). We split the dataset into training and test sets, in a proportion of 80-20% of items. Table 4 shows the format of the training set. The test set has been provided without gold labels (i.e. without the "Event" attribute) for testing purposes.

3.4 Data release

Both the training and the test sets were released on our website and protected with a password. As described in Section 3.3, the development data con-

File	Engagement	Date	Manip.	Visual	Text	Meme
1.jpg	21,053	22/08/19	1	Conte	aiuto	0
56.jpg	114	22/08/19	0	Salvini	alle solite	1

Table 2: An excerpt from the dataset for Task A, Meme Detection.

File	Engagement	Manip.	Visual	Text	Hate Speech
62.jpg	21,053	1	Conte	aiuto	0
114.jpg	12,572	1	Salvini	merdman	1

Table 3: An excerpt from the dataset for Task B, Hate Speech Identification.

File	Engagement	Date	Macro	Manip.	Visual	Text	Event
43.jpg	21,053	22/08/19	1	1	Conte	aiuto	1
23.jpg	114	22/08/19	1	0	Salvini	alle solite	0
114.jpg	12,572	25/08/19	0	1	Salvini	merdman	2

Table 4: An excerpt from the dataset for Task C, Event Clustering.

Team Name	Affiliation	Task
DMT	RN Podar School	A
Keila	Dipartimento di Matematica e Informatica di Perugia	A
UniTor	Università degli Studi di Roma "Tor Vergata"	A,B,C
UPB	Univesity Politehnica of Bucharest	A,B
SNK	ETI3	A

Table 5: Participants along with their affiliations and the tasks they participated in.

sisted of three distinct datasets, one for each task. The participants could download a distinct folder for each task, which contained:

- A UTF-8 encoded comma separated ".csv" file with 800 items (1,600 for task A), containing the metadata described in Section 3.3;

- A folder containing the images in .jpg format;

- A .csv file containing the relative image embeddings.

As for the test data, we released three folders whose structure is similar to the ones of the training sets. Each folder for the train sets contains:

- A UTF-8 encoded comma separated ".csv" file with 200 items (400 for Task A), which features the same metadata of the corresponding training set minus the golden label (i.e. "Meme" for Task A, "Hate speech" for Task B and "Event" for Task C);

- A folder containing the images in .jpg format;

- A .csv file containing the relative image embeddings.

All material was released for non-commercial research purposes only under a Creative Common license (BY-NC-ND 4.0). Any use for statistical, propagandistic or advertising purposes of any kind is prohibited. It is not possible to modify, alter or enrich the data provided for the purposes of redistribution.

4 Evaluation Measures

For all tasks, the models have been evaluated with $Precision$, $Recall$, and F_1 scores defined as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where TP are true positives, and FN and FP are false negatives and false positives, respectively. We computed $Precision$, $Recall$, and F_1

Figure 2: Two examples from the dataset for Hate Speech Identification: the meme at the top is classified as hate speech content, whereas the meme at the bottom is not.

for Task A and Task B considering only the positive class. For what concerns Task C, which is a multiclass classification task, we computed the performance for each class and then calculated the macro-average over all classes.

Different baselines were used for the different tasks:

Task A: Meme Detection. The baseline is given by the performance of a random classifier, which labels 50% of images as meme.

Task B: Hate Speech Identification. The baseline is given by the performance of a classifier labeling a meme as offensive when the meme text

contains at least a swear word[1].

Task C: Event Clustering. The baseline is given by the performance of a classifier labeling every meme as belonging to the most numerous class (i.e. the residual one).

5 Participants and Results

In total, 16 teams registered for DANKMEMES, and five of them participated in at least one of the tasks: DankMemesTeam (DMT) (Setpal and Sarti, 2020), Keila, UPB (Vlad et al., 2020), SNK (Fiorucci, 2020), and UniTor (Breazzano et al., 2020).

All of the 5 teams participated in Task A, while 2 teams participated in Task B and 1 in Task C. Participants could submit up to two runs per task: all of the teams did so consistently across tasks, with the exception of one team submitting a single run in Task A. This amounts to 9 runs for Task A, 4 for Task B and 2 for Task C, as detailed in Table 5.

Task A: Meme Detection. Task A consisted in differentiating between a meme and a not-meme. Five teams presented a total of 9 runs, as detailed in Table 6. The best scores have been achieved by the UniTor team with an F_1-measure of 0.8501 (with a Precision score of 0.8522 and a Recall measure of 0.848). The SNK and UPB teams followed closely, but all teams consistently showed a drastic improvement over the baseline.

Team	Run	Recall	Precision	F_1
Unitor	2	0.8522	0.848	0.8501
SNK	1	0.8515	0.8431	0.8473
UPB	2	0.8543	0.8333	0.8437
Unitor	1	0.839	0.8431	0.8411
SNK	2	0.8317	0.848	0.8398
UPB	1	0.861	0.7892	0.8235
DMT	1	0.8249	0.7157	07664
Keila	1	0.8121	0.6569	0.7263
Keila	2	0.7389	0.652	0.6927
baseline	1	0.525	0.5147	0.5198

Table 6: Results of Task A.

Task B: Hate Speech Identification. Task B consisted in the identification of whether a meme

---

[1]The list of swear words was downloaded from: `https://www.freewebheaders.com/italian-bad-words-list-and-swear-words/` (last access: 2nd November 2020).

Figure 3: Examples of memes from the dataset for Event Clustering task. Each meme refers to an event: (a) Beginning of the governement crisis; (b) Conte's speech and beginning of consultations; (c) Conte is called to form a new government; (d) 5SM holds a vote on the platform Rousseau.

is offensive or not. As detailed in Table 7, 2 teams participated in this task for a total of 4 runs (2 each). The best scores are achieved by the UniTor team for the F_1-measure at 0.823 and the Recall score of 0.8667, while the UPB team scored the best Precision measure at 0.8056. The scores improve over the baseline consistently across teams for what concerns the Recall score and the F_1-measure, while the Precision measure was not reached by any participant.

Team	Run	Recall	Precision	F_1
UniTor	2	0.7845	0.8667	0.8235
UniTor	1	0.7686	0.8857	0.823
UPB	1	0.8056	0.8286	0.8169
UPB	2	0.8333	0.7143	0.7692
baseline	1	0.8958	0.4095	0.5621

Table 7: Results of Task B.

Task C: Event Clustering. Task C consisted in clustering memes into 5 events using supervised classification. As seen in Table 8, a single team participated with 2 runs: the best score is therefore that of the UniTor team, with an F_1-score of 0.2657.

Team	Run	Recall	Precision	F_1
UniTor	1	0.2683	0.2851	0.2657
UniTor	2	0.2096	0.2548	0.2183
baseline	1	0.096	0.2	0.1297

Table 8: Results of Task C.

6 Discussion

We compare the participating systems according to the following main dimensions: classifi-

cation framework, exploitation of available features, multimodality of the adopted approaches, exploitation of further annotated data, and use of external resources. Since this is the first task about memes within the EVALITA campaign, we could not compare the obtained results with those achieved in any previous edition. A task about memes, Memotion, has been organized under SemEval 2020 (Sharma et al., 2020). However, the Memotion subtasks (Sentiment Classification, Humor Classification, and Scales of Semantic Classes) are quite different from those presented in DANKMEMES, and the results are hardly comparable.

System architecture. All the submitted runs to DANKMEMES leverage on neural networks, including very simple but equally efficient architectures. Multi-Layer Perceptrons (MLP) have been adopted by UniTor and SNK, ranked first and second in the the Meme Detection task, respectively. UPB adopted a Vocabulary Graph Convolutional Network (VGCN) combined with BERT contextual embeddings for text analysis. This team employed this architectural design within a Multi-Task Learning (MTL) technique, based on two main neural network components: one for the text and the other for the image analysis. The outputs of these two elements were concatenated and used to feed a Dense layer. The system in DMT is composed of three 8-layer feed-forward networks, each taking as input a different image vector representation. Finally, Keila exploited Convolutional Neural Networks (CNN) in each of the submitted run.

External resources. All the presented models employed external resources to feed their neural architecture with image and text representations. The text contained in the images was encoded by using different flavours of word embeddings. Most of the participants exploited one of the available BERT contextual embeddings model for the Italian language (AlBERTo, UmBERTo, or GilBERTo). However, with its first run, SNK achieved the second position in the Meme Detection task using the pre-trained FastText embeddings for the Italian language. Similarly, Keila adopted pre-trained Word2Vec for the Italian language, though achieving lower results. As for the visual channel, the DANKMEMES datasets provided a state-of-the-art representation of images, obtained with the ResNet50 architecture. Most of the participants experimented the use of other image vector representations as well: DMT used three different image vector: AlexNet, ResNet, and DenseNet; UniTor and UPB examined several models, among which: EfficientNET, VGG-16, YOLOv4, ResNet50, and ResNet152. UniTor chose EfficientNet for their final models, while UPB based their ssystems on ResNet50 and ResNet152.

Multimodality. The exploitation of both images and text turned out to be fundamental for the task of Meme Detection. Since memes adhere to specific visual conventions, participants tried to exploit visual data at their best. The first run of UniTor only relied on an image classifier, whereas DMT exploited the information resulting from three different image classification models, then combined with word embeddings. Nevertheless, the best results were obtained by the combination of text and image information. In its second run, UniTor concatenated the image representation returned by their first model with pre-trained contextual word embeddings fine-tuned on DANKMEMES data. Similarly, SNK and UPB leveraged both textual and image data. Keila was the only participant who did not combine text and image information in any of the submitted runs. For what concerns the second task, the first UniTor run only relied on textual data and was slightly overcame only by their second run. As observed by the team, in the Hate Speech Identification task, textual data heavily impact the classification results. Finally, UPB combined both image and textual data for this task.

Data Augmentation. Several participants chose to adopt a data augmentation technique. UniTor successfully manipulated the provided images by horizontally mirroring them. On the contrary, DMT created nine versions of each image at first, editing brightness, rotation, and zoom, but then dropped them due to the overfitting caused by the unmodified metadata associated with each image. Keila augmented textual data by firstly translating the image texts in English and then back to Italian. Regarding the second task on Hate Speech Identification, UniTor trained for a few epochs the UmBERTo embeddings on a dataset made available within the Hate Speech Detection (HaSpeeDe) task (Bosco et al., 2018) before training it on the DANKMEMES dataset.

Exploited features. SNK encoded and concatenated in a single vector picture manipulation, visual, and engagement, along with the sentence and the image representation of each meme. Keila employed engagement and manipulation features as well. DMT normalized engagement and represented dates with the count of days from a selected reference date. Along with the other provided data, temporal features were exploited by UPB as well, through the computation of complementary sine and cosine distances, in order to preserve the cyclic characteristics of days and months. Finally, UniTor relied only on visual and textual information.

Event Clustering. The goal of this task was to assign each meme to the event it refers to. Only UniTor participated in this task, modeling it as a classification problem in two distinguished runs. The first model only exploited textual data representation provided by the Transformer architecture to feed the MLP classifier. Furthermore, UniTor submitted a second run. The team mapped the original classification problem, which counted five different labels (each corresponding to an event) over a binary classification one. After pairing a meme to each event, a pair was labeled as positive if the association was correct, negative otherwise. However, this run did not overpass the first one, the outcome of which doubled the provided baseline.

7 Final Remarks

The paper describes a task for the detection and analysis of memes in the Italian language.

DANKMEMES is the first task of this kind in the EVALITA campaign. Although memes are widespread on the Web, it is still hard to define them precisely. However, DANKMEMES highlighted the fundamental role of multimodality in memes detection, mainly the combined use of texts and images for their classification. Therefore, we could say that memes share peculiar linguistic features, other than conventional layouts. Future work will focus on the extension of the dataset, which showed some limitations, especially for its reduced size and for the unbalanced representation of some events. This is due to the difficulty of meme collection, especially when filtered in relation to a specific event (e.g., the 2019 Italian government crisis).

References

Eisa Al Nashmi. 2018. From selfies to media events: How instagram users interrupted their routines after the charlie hebdo shootings. *Digital Journalism*, 6(1):98–117.

Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro. 2020. Evalita 2020: Overview of the 7th evaluation campaign of natural language processing and speech tools for italian. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.

Cristina Bosco, Dell'Orletta Felice, Fabio Poletto, Manuela Sanguinetti, and Tesconi Maurizio. 2018. Overview of the evalita 2018 hate speech detection task. In *EVALITA 2018-Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*, pages 1–9.

Claudia Breazzano, Edoardo Rubino, Danilo Croce, and Roberto Basili. 2020. Unitor @ dankmemes: Combining convolutional models and transformer-based architectures for accurate meme management. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.

Patrick Davison. 2012. The language of internet memes. *The Social Media Reader*, pages 120–134.

Richard Dawkins. 2016. *The Selfish Gene*. Oxford University Press.

Stefano Fiorucci. 2020. Snk @ dankmemes: Leveraging pretrained embeddings for multimodal meme detection. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.

Jean French. 2017. Image-based memes as sentiment predictors. In *2017 International Conference on Information Society (i-Society)*.

Noam Gal, Limor Shifman, and Zohar Kampf. 2016. "it gets better": Internet memes and the construction of collective identity. *New media & society*, 18(8):1698–1714.

Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Giulia Giorgi and Ilir Rama. 2019. "one does not simply meme". framing the 2019 italian government crisis through memes. In *La comunicazione politica nell'ecosistema dei media digitali Convegno dell'Associazione Italiana di Comunicazione Politica (ASSOCOMPOL)*.

Hugo Gonçalo Oliveira, Diogo Costa, and Alexandre Pinto. 2016. One does not simply produce funny memes! – explorations on the automatic generation of internet humor. In *Proceedings of the Seventh International Conference on Computational Creativity (ICCC 2016)*.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Douwe Kiela and Léon Bottou. 2014. Learning image embeddings using convolutional neural networks for improved multi-modal semantics. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 36–45.

Michele Knobel and Colin Lankshear. 2007. Online memes, affinities, and cultural production. *A new literacies sampler*, 29:199–227.

Klaus Krippendorff. 2018. *Content analysis: An introduction to its methodology*. Sage publications.

Ryan M. Milner. 2016. *The World Made Meme: Public Conversations and Participatory Media*. MIT Press.

Abel L. Peirson V and E. Meltem Tolunay. 2018. Dank learning: Generating memes using deep neural networks. *CoRR*, abs/1806.04510.

Vasiliki Plevriti. 2014. Satirical user-generated memes as an effective source of political criticism, extending debate and enhancing civic engagement.

Andrew S. Ross and Damian J. Rivers. 2017. Digital cultures of political participation: Internet memes and the discursive felegitimization of the 2016 us presidential candidates. *Discourse, Context & Media*, 16:1–11.

Benet Oriol Sabat, Cristian Canton Ferrer, and Xavier Giro-i Nieto. 2019. Hate speech in pixels: Detection of offensive memes towards automatic moderation. *arXiv preprint arXiv:1910.02334*.

Jinen Setpal and Gabriele Sarti. 2020. Dankmemesteam @ dankmemes: Archimede: A new model architecture for meme detection. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.

Chhavi Sharma, Deepesh Bhageria, William Scott, Srinivas PYKL, Amitava Das, Tanmoy Chakraborty, Viswanath Pulabaigari, and Bjorn Gamback. 2020. Semeval-2020 task 8: Memotion analysis – the visuo-lingual metaphor! *arXiv preprint arXiv:2008.03781*.

Limor Shifman. 2013. Memes in a digital world: Reconciling with a conceptual troublemaker. *Journal of computer-mediated communication*, 18(3):362–377.

E. S. Smitha, Selvaraju Sendhilkumar, and G. S. Mahalaksmi. 2018. Meme classification using textual and visual features. In *Computational Vision and Bio Inspired Computing*, pages 1015–1031.

Emi Tanaka, Timothy Bailey, and Uri Keich. 2014. Improving meme via a two-tiered significance analysis. *Bioinformatics*, 30:1965–1973, 03.

George-Alexandru Vlad, George-Eduard Zaharia, Dumitru-Clementin Cercel, and Mihai Dascalu. 2020. Upb @ dankmemes: Italian memes analysis: Employing visual models and graph convolutional networks for meme identification and hate speech detection. In Valerio Basile, Danilo Croce, Maria Di Maro, and Lucia C. Passaro, editors, *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)*, Online. CEUR.org.

Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar. 2019a. Predicting the type and target of offensive posts in social media. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1415–1420.

Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar. 2019b. SemEval-2019 task 6: Identifying and categorizing offensive language in social media (OffensEval). In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 75–86.