

## Instruction for Participants:

# SardiStance @ EVALITA2020

**Mirko Lai**

Dipartimento di Informatica  
Università degli Studi di Torino  
lai@di.unito.it

**Alessandra Teresa Cignarella**

Università degli Studi di Torino  
Universitat Politècnica de València  
cigna@di.unito.it

**Cristina Bosco**

Dipartimento di Informatica  
Università degli Studi di Torino  
bosco@di.unito.it

**Viviana Patti**

Dipartimento di Informatica  
Università degli Studi di Torino  
patti@di.unito.it

**Paolo Rosso**

Departamento de Sistemas Informáticos y Computación  
Universitat Politècnica de València  
proso@dsic.upv.es

## Contents

<b>1</b>	<b>Task description</b>	<b>2</b>
<b>2</b>	<b>Training and Test Data</b>	<b>3</b>
2.1	Format and Distribution . . . . .	4
<b>3</b>	<b>Machine Learning for Beginners</b>	<b>5</b>
<b>4</b>	<b>How to submit your runs</b>	<b>5</b>
<b>5</b>	<b>Submission format</b>	<b>6</b>
<b>6</b>	<b>Evaluation</b>	<b>6</b>
6.1	Baselines . . . . .	7
<b>7</b>	<b>Final remarks</b>	<b>7</b>

# 1 Task description

With this task we invite participants to explore features based on the textual content of the tweet, such as structural, stylistic, and affective features, but also features based on contextual information that does not emerge directly from the text, such as for instance knowledge about the domain of the political debate or information about the user's community. Overall, we propose two different subtasks:

- **Task A - Textual Stance Detection:**

The first task is a three-class classification task where the system has to predict whether a tweet is in FAVOUR, AGAINST or NEUTRAL/NONE towards the given target (following the guidelines below), exploiting only textual information, i.e. the text of the tweet.

---

From reading the tweet, which of the options below is most likely to be true about the tweeter's stance or outlook towards the target?

1. **FAVOUR:** We can infer from the tweet that the tweeter supports the target.
2. **AGAINST:** We can infer from the tweet that the tweeter is against the target.
3. **NEUTRAL/NONE:** We can infer from the tweet that the tweeter has a neutral stance towards the target or there is no clue in the tweet to reveal the stance of the tweeter towards the target (support/against/neutral) [2].

---

- **Task B - Contextual Stance Detection:**

The second task is the same as the first one: a three-class classification task where the system has to predict whether a tweet is in FAVOUR, AGAINST or NEUTRAL/NONE towards the given target. Here participants will have access to a wider range of contextual information based on:

- the post such as: the number of retweets, the number of favours, the number of replies and the number of quotes received to the tweet, the type of posting source (e.g. iOS or Android), and date of posting.
- the user, number of tweets ever posted, user's bio (only emojis), user's number of followers, user's number of friends.

- the social network, such as: friends, replies, retweets, and quotes' relations.  
The personal ids of the users will be anonymized but their network structures will be maintained intact.

Participants can decide to participate to both tasks or only to one. Although they are encouraged to participate to both.

**Examples** The following table show some examples of tweets annotated for the task.

tweet_id	tweet_text	label
1	LE SARDINE IN PIAZZA MAGGIORE NON SONO ITALIANI SE LO FOSSERO NON SI METTEREBBERO CONTRO LA DESTRA CHE AMA L'ITALIA E VUOLE RIMANERE ITALIANA	AGAINST
2	Non ci credo che stasera devo andare in teatro e non posso essere fra le #Sardine #Bologna #bolognanonsilega	FAVOUR
3	Mi sono svegliato nudo e triste perché a Bologna, tra salviniani e antisalviniani, non mi ha cagato nessuno.	NONE

Table 1: Some examples extracted from the dataset for the task.

## 2 Training and Test Data

We have collected tweets written in Italian about the *Sardines movement*<sup>1</sup>, retrieving tweets containing the keywords “sardina”, “sardine”, and the homonymous hashtags. Furthermore we have collected all the conversation threads connoted to the said tweet, iteratively following the reply’s tree. This enrichment has enlarged the dataset also with tweets that don’t actually contain the keywords used for retrieval. To extract a random sample for the manual annotation task, retweets, quotes and replies are have been discarded. We also discarded tweets that contain URLs or extended entities (videos, photos, GIFs etc..).

The entire “*StardiStance Dataset*” consists of 3,242 instances. As usual in evaluation contests, data are split in two sets: the first one for development and training, the other for testing participant’s results. The former includes 2,132 tweets while the latter is composed of the remaining 1,110.

The training set is made available on **May 29th, 2020**. It is associated with a Creative Commons 4.0 license<sup>2</sup> that defines the terms that the users will be bind by in data exploitation and citation.

If some **update** is applied to the released data, a notification will be published in the task website<sup>3</sup> and on the Google Group<sup>4</sup>.

The test set will be instead released on **September 4th, 2020** (see also the “Important dates” section in the task web page).

Each participating team will initially have access to the training data only. Later, the unlabeled test data will be released. After the assessment, the complete test data will be released as well.

<sup>1</sup>[https://en.wikipedia.org/wiki/Sardines\\_movement](https://en.wikipedia.org/wiki/Sardines_movement).

<sup>2</sup><https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>

<sup>3</sup><http://www.di.unito.it/~tutreeb/sardistance-evalita2020/index.html>

<sup>4</sup>Join the Google Group <https://groups.google.com/forum/#!forum/sardistance-evalita2020> and check it regularly to be sure that you updated about news regarding the task.

## 2.1 Format and Distribution

A single training set will be provided for both the tasks A and B, which includes 2,132 tweets distributed as in Table 2.

TRAINING SET			
AGAINST	FAVOR	NEUTRAL	TOTAL
1,028	589	515	2,132

Table 2: Distribution of Data

The training set will be made available at the following URL on May 29th, 2020:

<https://github.com/mirkolai/evalita-sardistance>

The dataset is password protected. In order to obtain the password you need to register to the Google Group “*SARDISTANCE @ EVALITA2020*” and to sign the Copyright Agreement by compiling this form: <https://forms.gle/xuikYEsHB18uVVQ67>.

### Task A

The training data (TRAIN.csv) is released in the following format:

```
tweet_id  user_id  text  label
```

where **tweet\_id** is the Twitter ID of the message, **user\_id** is the Twitter ID of the user who posted the message, **text** is the content of the message, **label** is AGAINST, FAVOR or NONE.

### Task B

In order to participate to Task B, exploiting also contextual information regarding tweet, user and social network we release some additional data.

- In the file TWEET.csv, containing contextual information regarding the tweet you will find the following format:

```
tweet_id  user_id  retweet_count  favorite_count  source  created_at
```

where **tweet\_id** is the Twitter ID of the message, **user\_id** is the Twitter ID of the user who posted the message, **retweet\_count** indicates the number of times the tweet has been retweeted, **favorite\_count** indicates the number of times the tweet has been liked, **source** indicates the type of posting source (e.g. iOS or Android), and **created\_at** displays the time of creation according to a yyyy-mm-dd hh:mm:ss format. Minutes and seconds have been encrypted and transformed to zeroes for privacy issues.

- The file USER.csv contains contextual information regarding the user. It is released in the following format:

`user_id statuses_count friends_count followers_count created_at emoji`

where `user_id` is the Twitter ID of the user who posted the message, `statuses_count`, `friends_count` indicates the number of friends of the user, `followers_count` indicates the number of followers of the user, `created_at` displays the time of the user registration on Twitter, and `emoji` shows a list of the emojis in the user's bio (if present, otherwise the field is left empty).

- The files `FRIEND.csv`, `QUOTE.csv`, `REPLY.csv` and `RETWEET.csv` contain contextual info about the social network of the user. Each file is released in the following format:

`Source Target Weight`

where `Source` and `Target` indicate two nodes of a social interaction between two Twitter users. More specifically, the source user performs one of the considered social relation towards the target user. Two users are tied by a friend relationship if the source user follows the target user (friend relationship does not have a weight, because it is either present or absent); while two users are tied by a quote, retweet, or reply relationship if the source user respectively quoted, retweeted, or replied the target user.

`Weight` indicates the number of interactions existing between two users. Note that this information is not available for the friend relation (hence, this column is not present in the `FRIEND.csv` file) due to the fact that it is a relationship of the type present or absent and cannot be described through a weight. In all the files, users are defined by their anonymized Twitter ID.

### 3 Machine Learning for Beginners

In order to encourage task participation from students of all ages and everyone who is curious towards the fields of Machine Learning and Natural Language Processing, we have developed a simple automatic system that can be exploited by everyone as a starting point to participate to our task.

You can find it here: <https://github.com/mirkolai/evalita-sardistance/tree/master/machine%20learning>.

### 4 How to submit your runs

Once you have run your system over the test data that you have downloaded, you will have to send it to us following these recommendations:

1. Choose a team name and name the files containing your runs in the following way:
  - `sardistance2020_TeamName_RunNumber_c` for the constrained runs

- `sardistance2020_TeamName_RunNumber_u` for the unconstrained runs
2. Send all relevant files to the following address: `mirko.lai@unito.it` using the subject “sardistance2020 – TeamName”.
  3. In the body of the email please specify all resources you used in the unconstrained run (if any).
  4. Please note that the test set consists of 1,110 tweets, so you can double check that your files are complete by verifying that you have the correct number of lines.

## 5 Submission format

Results for all tasks should be submitted in a plain text file with tab-separated fields (tsv). The format of the run files submitted by participants is the same for both tasks:

```
"tweet_id"  "label"
```

Each submitted run must contain one tweet per line, including the `tweet_id` and the predicted label.

**Number and type of runs** For each task, we distinguish between constrained and unconstrained runs:

- for a constrained run, teams must use only the provided training data; other resources, such as lexicons are allowed; however, it is not allowed to use additional training data in the form of tweets or sentences with annotations for stance;
- for an unconstrained run, teams can use additional data for training, e.g., additional tweets annotated for stance (and it must be specified within the submission of the runs).

Please note that we will provide two separate ranks for the constrained and unconstrained runs. **Important:** if you take part in a given task, you must submit at least a constrained run, while the unconstrained one is optional. Each team may perform up to two submissions for the constrained run. Similarly, we allow a maximum of two submissions for the unconstrained run per team. Participants are invited to submit multiple runs to experiment with different models and architectures, but discouraged from submitting slight variations of the same model.

## 6 Evaluation

We will provide a separate official ranking for Task A and Task B, and two separate ranking for constrained and unconstrained runs. Systems will be evaluated using F1-score computed over the two main classes (FAVOUR and AGAINST). The submissions will be ranked by the averaged F1-score above the two classes.

We will measure the accuracy of the precision, the precision, recall and F-score for all three classes:

## 6.1 Baselines

For both sub-tasks, we will compute a first baseline using a simple machine learning model based on SVM combined with uni-gram feature. And as second baseline we will provide one measure computed by a model based on previous work from the authors (MultiTacos) [1].

## 7 Final remarks

If you have any questions or problems, please start a topic on the Google Group: [sardistance-evalita2020@googlegroups.com](mailto:sardistance-evalita2020@googlegroups.com) or check for more info on the SardiStance @ EVALITA 2020 official Web Page: <http://www.di.unito.it/sardistance-evalita2020.html>.

## References

- [1] LAI, M., CIGNARELLA, A. T., HERNANDEZ FARIÁS, D. I., BOSCO, C., PATTI, V., AND ROSSO, P. Multilingual Stance Detection in Social Media: Four Political Debates on Twitter. *Computer Speech and Language* 63 (2020).
- [2] MOHAMMAD, S., KIRITCHENKO, S., SOBHANI, P., ZHU, X., AND CHERRY, C. A Dataset for Detecting Stance in Tweets. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (Paris, France, may 2016), N. C. C. Chair), K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds., European Language Resources Association (ELRA), pp. 3945–3952.