Sentiment prediction of Twitter contents

Edoardo Chiò

Politecnico di Torino

Student id: s301486

s301486@studenti.polito.it

Abstract—**The abstract goes here. Keep it short (approx. 3-4 sentences)**

I. PROBLEM OVERVIEW

This project concerns a classification problem applied to a collection of Twitter posts (i.e., *tweets*) written by different users. The goal of the project is to perform a sentiment analysis of the posts contained in the dataset.

Training and validation of the models are conducted on a development set, containing 224,994 labelled recordings, while the set to test the models contains 74,999 recordings.

The development set is composed of six fields:

- *sentiment*: field of the sentiment labels; a label is associated to each record, having value **1** if the text has a positive trait, **0** otherwise;
- *ids*: numerical identifier of the tweet;
- *date*: publication date of the tweet;
- *flag*: query used to collect the tweet;
- *user*: name of the user that posted the tweet;
- *text*: text of the tweet.

Analysing the development set it is possible to notice that it is not well balanced: 130,157

II. PROPOSED APPROACH

In this section, you will present your solution. Please fill in accordingly.

You can use citations as follows: [1] (you can add BibTeX citations in the *bibliography.bib* file).

A. Preprocessing

B. Model selection

C. Hyperparameters tuning

III. RESULTS

Here you will present your results (models & configurations selected, performance achieved)

IV. DISCUSSION

Any relevant discussion goes here.

REFERENCES

[1] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.