

[F23] PracticalMLandDL

Project Intermediate Report #1

1. **Roman Makarov**

o.makarov@innopolis.university

2. **Adela Krylova**

a.krylova@innopolis.university

Topic: **Real or Not? NLP with Disaster Tweets**

GitHub: https://github.com/rmakarovv/pmldl_project

Collab:

https://colab.research.google.com/drive/1aJqS3D5_N0tBp-pNsrFhagknLDN_MK7IR?usp=sharing

Intermediate progress:

- We revised NLP and ML techniques
 - We found and downloaded appropriate data
 - We made preliminary analysis of the data
 - We suggested several possible techniques for solving the task
 - We planned our future work
-

Preliminary results:

- We analysed the data distribution (frequency of words, amount of data per class, number of NaN data...)
 - We did some preliminary data preprocessing (stop words and punctuation removal, count vectorization)
 - We suggested further data preprocessing techniques - for example: embeddings, lemma extraction...
 - We suggested possible solutions worth trying, for example: NLP based word frequency, Logistic regression, Bidirectional LSTM, NN...
-

Tasks distribution:

So far, we worked together on all the steps. During the next weeks we plan to work individually, each of us trying to apply different techniques for solving our task and compare our results.