Your Tech Bootcamp

Group # - Every Second Matters: 🚨an app for timely disaster response🚨

Group members

Yağmur Helin Aslan #promo: fs-pt-08

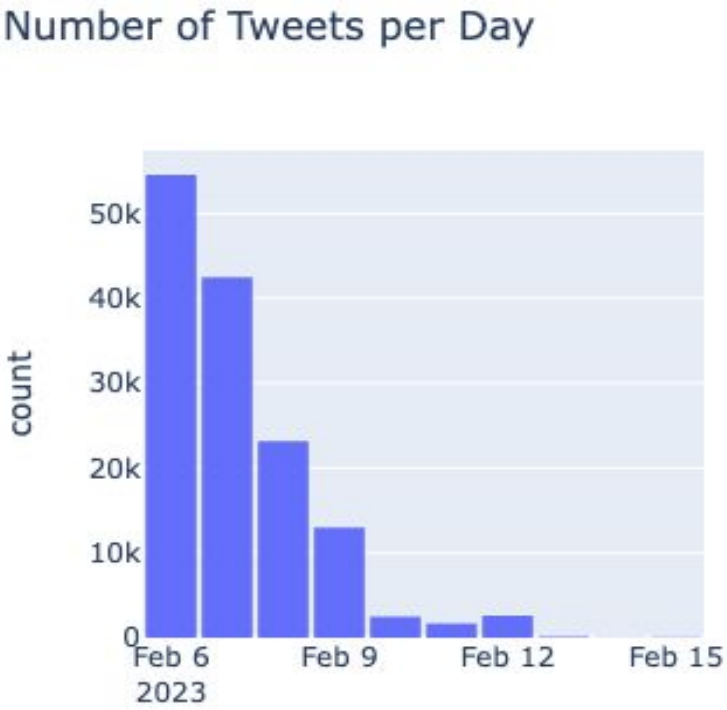
M2 student in Economie du Développement Durable, Paris 1 Panthéon-Sorbonne

Linkedin: www.linkedin.com/in/yagmur-helin-aslan

Pitch

Using Turkish twitter data from the earthquake that hits Syria and Turkey on 6 February, this app will perform three task:

- Detection of rescue calls and emergency needs with a binary text classifier
- Extraction of person names, city names and addresses in emergency tweets with a Named Entity Recognition (NER) model
- Geoplotting of those addresses on an interactive map simulating a real-time plot



Text Annotation

Other 0 | Rescue_call 1 | Urgent_need 2

Turunçlu mahallesi samandag yolu uzeri, saray market yanı 95/B Defne-Hatay Enkazda kalanlardan biri Nilay Oltacı İletişim 05161646506 #Turkey #CristianoRonaldo #hatayyardimbekliyor #hatayiskenderun #HalkTV #özgürdemirtaş #fulyaöztürk #EnkazAltında #tahaduymaz

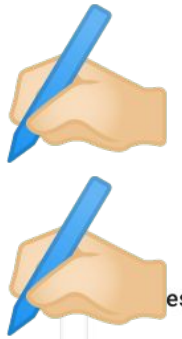
Odabasi Mahallesi. Ali gaffar okan bulvari no:12 dağsu Apartmanı Seda Aksu Mustafa
•ADDR •PER •PER

Aksu Seda hamile acil yardım Antakya/Hatay Mehmet Aksu (Mustafa Aksu'nun
•ADDR •CITY •PER •PER

kardeşi) Tel:05350815624 #Hatay #hatayyardimbekliyor #earthquake @ahbap
•NUM

@BabalaTv @haluklevent @KendineMuzisyen @OguzhanUgur

https://t.co/nKOVykp29E



ess

Total 2304

Complete 1352

59%

Label Types

PER p | ADDR a | CITY c | NUM n

Task 1: Binary Text Classification

❓Text Preprocessing

Manually built processor:

- remove non-alphanumeric characters (incl. emojis)
- recognize special turkish characters (çÇğĞıİöÖşŞüÜ)
- remove all entities that follow tags(@) and hashtags(#)
- lowercasing

Zemberek NLP for preprocesing:

- sentence normalization by fixing typos,
- normalizing tr-en keyboard differences

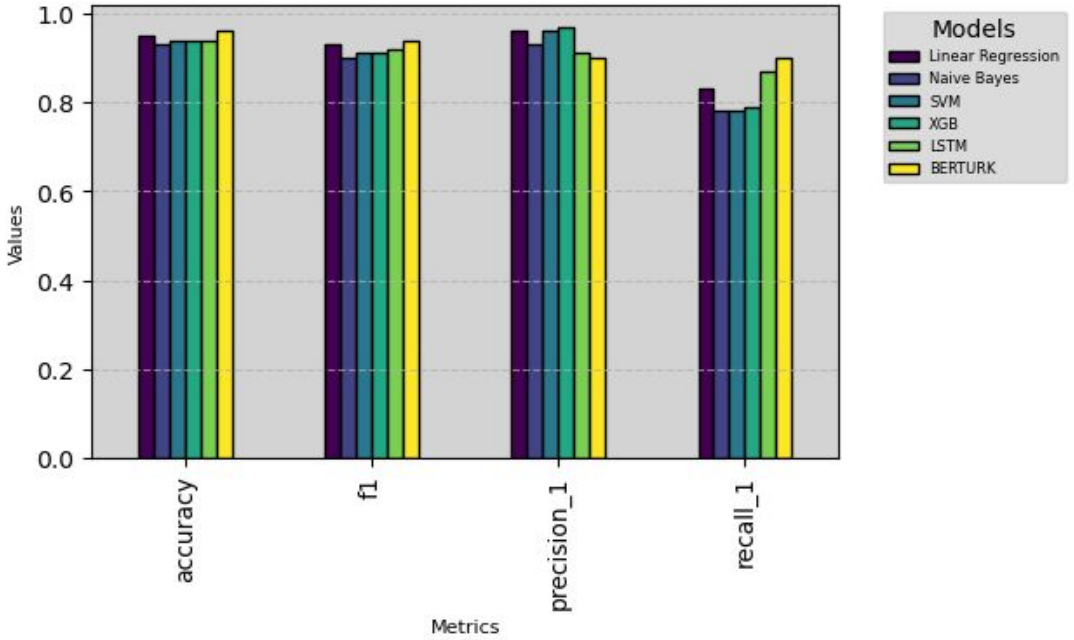
…or None? 😈

Model Comparison : Choosing right metrics

K-fold cross-validation results:

processor	model	accuracy	f1
Manual	Logistic Regression	0.93	0.85
	Naive Bayes	0.92	0.84
	Linear SVM	0.92	0.84
	Kernel SVM	0.93	0.84
Zemberek	Logistic Regression	0.92	0.85
	Naive Bayes	0.92	0.84
	Linear SVM	0.93	0.85
	Kernel SVM	0.93	0.84

Performance metrics:



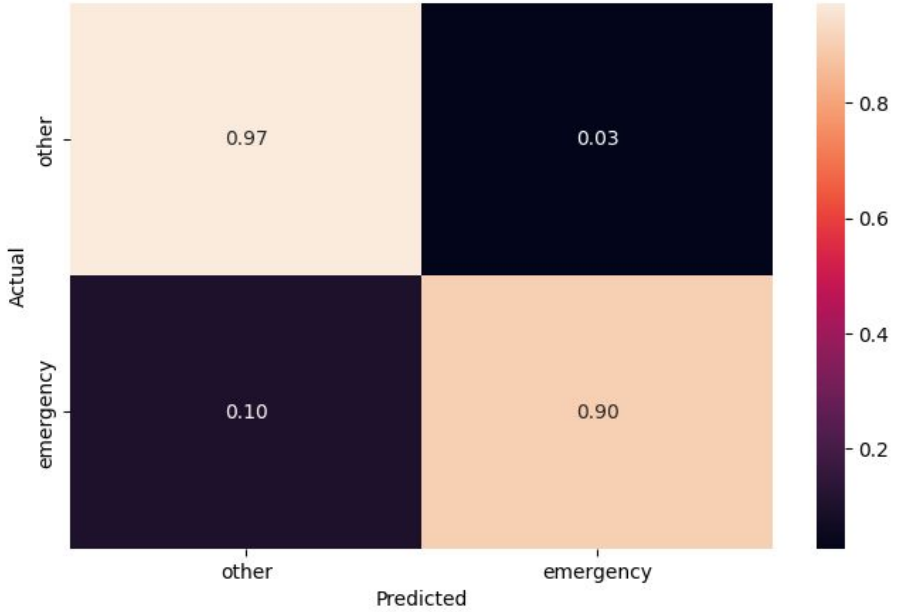
BERT-based: Transfer learning and Fine-tuning

Check this out:

- LABEL_1 : emergency
- LABEL_0 : other

Arkadaşım Emine DİLKEN ve ailesine ulaşamıyoruz lütfen yardım edin. Bulundukları konuma yardım ulaşmamış. Çalışmaları bırakmışlar Hamidiye sitesi, Şehit Abdullah Çavuş, 46050 Onikişubat/Kahramanmaraş #hamidiyesitesiyardım #deprem #YARDIM #Turkey @ahbap'

('We cannot reach my friend Emine DİLKEN and her family, please help. No help has reached their location. They left the works Hamidiye site, Şehit Abdullah Çavuş, 46050 Onikişubat/Kahramanmaraş #hamidiyesitesiyardım #deprem #earthquake #YARD #Turkey @ahbap')



Task 2: Named Entity Recognition

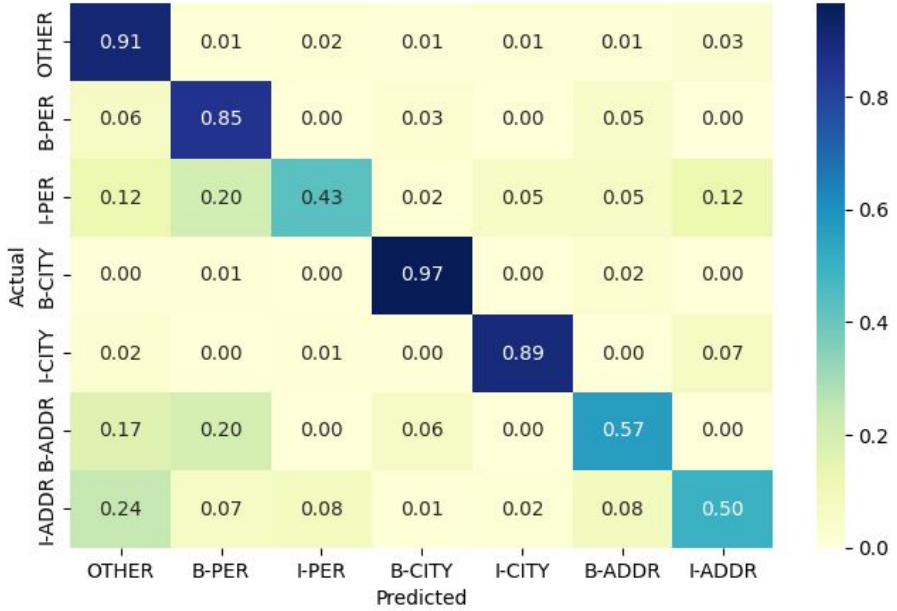
['O', 'B-PER', 'I-PER', 'B-ORG', 'I-ORG', 'B-LOC', 'I-LOC'] ➡️ *["O", 'B-PER', 'I-PER', 'B-CITY', 'I-CITY', 'B-ADDR', 'I-ADDR']*

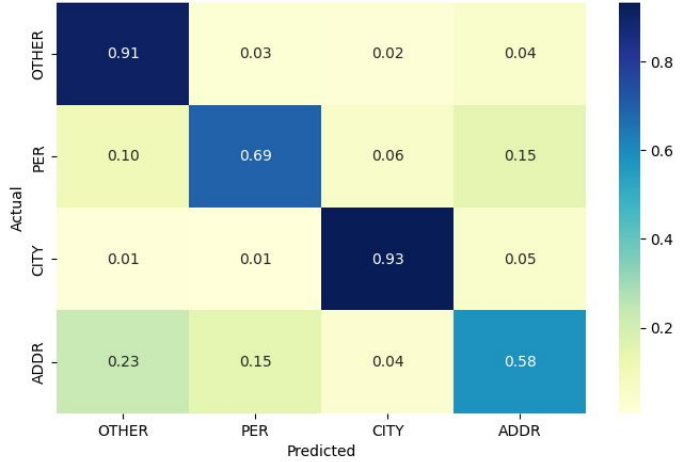
❓❓❓ **Text Preprocessing :**

- tons of abbreviations (*apt, sk /sok, cd /cad, mah /mh…*)
- variations in city names

‘Kahramanmaraş’:
KMaraş, K.Maraş,
Maraş, k.maras, Maras,
Kahramanmaras,
K.maraş, maraş, …

Metrics Evaluation





Check it out with an example tweet :

“Turgut Reis mahallesi Zey Caddesi No - 40 Adıyaman Merkez Zombaba Cami Karşısı Polis Memuru Fatih Sarıgül Saatlerdir haber alınamıyor, gören duyan bilen Lütfen Yardımcı olsun #deprem #turkey”

Task 3: Geocoding and Plotting

- Google Maps API to geocode addresses in latitudes and longtitudes
- pyplot scatter_mapbox to visualize on a map

What’s next?

- **Caveats and challenges:** Twitter API, GPU limits, GPU reliance
- **Deployment**
- **Improvement and ideas:**
  - clarifying better the app’s purpose (extracting and mapping exact locations vs catching any emergency help call),
  - a less memory consuming and accelerated model
  - detecting status (missing person, under the rubbles, etc),

Jedha

Any questions ?