

## **All Data collected from Social Media Posts on Twitter, Facebook, Stormfront**

### **Example German Tweets**

#### **Positive**

Manche Kommentare hier sind einfach dumm! Ich kann es nicht anders beschreiben. Pegida ist radikal und die meisten Anhänger dümmliche Mitläufer! Wegen Menschen wie euch konnte der 2. Weltkrieg passieren. Euch werden Wörter wie Lügenpresse in den Mund gelegt und ihr plappert sie ohne zu denken nach. Ich schäme mich für euch!

Was ist mit der Vergewaltigung einer Schülerin in Gratz durch 5 Afrikaner...ich lese nichts davon in der Zeitung-mir ist es egal welcher Asylant in die Fresse kriegt,ich will diese elenden Schmarotzer nicht in Deutschland haben.Wir haben genug Arbeitslose...was wird für die getan.....

#### **Negative**

Die Nachrichten schau ich mir nicht mehr.

#### **Interesting examples which are not actually hate speech**

Deutschland ist vorbei! 30-50 Jahre , und es wird eine muslimische Mehrheit.

Ob #Asylanten wohl mitbekommen, daß #linke, #gutmenschen, #antifa, die Deutsche Nation, zerstören wollen? ?

Das Geld wird nicht für Flüchtlinge gebraucht das Geld soll nur die Löcher in den Rucksäcken der Politiker stopfen. Folge dem Geld das schlafschaf versteht es dann hoffentlich auch

mit leuten hier zu diskutieren die anderer meinung sind, um diese ins eigene boot zu holen, lohnt nicht. es gibt menschen die sehen die welt aus multi-kulti-agen, denen sind die wirtschaftliche last, die integrierbarkeit der immigranten, die folgen der überfremdung etc. völlig egal. es gibt aber auch menschen, die die heutigen probleme und auswirkungen einer überfremdung spürbar wahrnehmen, welche sich um die zukunft ihrer kinder und deutschlands sorgen, die die sozialen und wirtschaftlichen probleme und die ratlosigkeit der politiker europäischer länder über die zustände in den ghettos der europäischen großstädte mit offenen augen sehen. die probleme mit bestimmten einwanderungsgruppen sind europaweit identisch: hohe arbeitslosigkeit, kriminalität, geringe bildung etc. kurzum, wer sich nicht integrieren will, integriert sich nie. gesetze können keinen menschen zur integration zwingen, wenn die innere bereitschaft dazu fehlt. holland, das bis in die neunziger jahre einwanderungsfreundlichste land europas, hat alles versucht diese menschen zu integrieren und musste dennoch kapitulieren und als endgültige lösung die einwanderungsgesetze verschärfen. immigranten sind eben keine maschinen, die man durch das aufspielen von

integrationssoftware neu programmieren kann. im jahre 1970 gab es in deutschland 3 moscheen, heute sind es 3000. solange der wasserhahn für immigranten aus muslimischen ländern offen steht, ist die islamisierung nur eine frage der zeit. multi kulti funktioniert nicht einmal in dem einwanderungsland usa, aber dies erkennen verträumte multi kulti augen nicht. wer sich heute in seinen noch vorhandenen wohlstandskokon zurückzieht, handelt verantwortlos gegenüber seinen kindern und den nachfolgenden generationen. (CP)

## **How Training is Done ?**

[ If model is trained and tested on the same language then evaluation is referred to as in-domain evaluation while if training is done on English language and testing on another language then we term it as out-domain evaluation. ]

I use Cross Lingual Roberta Model for this task. This model is already unsupervised pretrained on a corpus curated from common crawl. The training strategy for the preraining is Masked Language Modelling and Next sentence prediction. This corpus has 104 languages.

### **I employ 4 strategies to adapt this model for the task of hate speech detection for multiple languages:**

1. In the first strategy the model is applied on out-domain example i.e ( German Post ) while only finetuned on English. This strategy is popularly called Zero-Shot evaluation.
2. In the second strategy I fine-tune the model on few in-domain examples i.e German posts ( supervised training) after fine-tuned on English posts and then evaluate on German Posts. This strategy is popularly called few shot evaluation.
3. In the third and the 4rth strategy before applying 1 & 2 I pretrain the model with unsupervised Masked Language objective on In-domain language.

In the initial pilot I show results with examples on three Indo-European Language families i.e Geman, Spanish, Italian

## GERMAN SOCIAL MEDIA POSTS

[ In All ZeroShot evaluation cases the model undergoes supervised finetuning on out-domain examples i.e English social media posts ]

### 1 ) ZeroShot Detection Examples [ AUCROC - 0.6962819343065694 ]

True Positive Examples:

Make Rape Legal"" findet Merkel auch und deshalb flutet sie unser Land mit #Rapefugees. #abmerkeln #merkelmussweg #merkeldiktatur

Kinderfickende, schächtende #rapefugees! Wer Kinder-, Frauen-, oder Tierrechte verteidigt, darf nicht #gruene #spd oder #cdu wählen! #ltwbw

False Positives Examples:

Das gerechte in #Deutschland ist ja, dass nicht nur #Juden, sondern auch #Christen verfolgt werden. #WirSchaffenDas #Merkel #Islamisierung'

False Negative Examples:

Die Alte Schabrake Namens Merkel muß Weg und ein Deutschland für Deutsche muß her!!!

### 2 ) FewShot detection Examples after Supervised fine tuning on 256 in-domain examples [ AUCROC- 0.8297552597681409 ]

Correct predicted example from Strategy 1

Die Alte Schabrake Namens Merkel muß Weg und ein Deutschland für Deutsche muß her!!!

Das gerechte in #Deutschland ist ja, dass nicht nur #Juden, sondern auch #Christen verfolgt werden. #WirSchaffenDas #Merkel #Islamisierung

### **3) ZeroShot Detection Examples after Unsupervised Pretraining on in-domain Examples [ AUCROC: 0.722131279519107]**

Correct example Predicted compared to strategy 1

Ja zur Einwanderung z.B. aus Amerika, Australien, Ostasien. Nein zu #Überfremdung und #Islamisierung.

### **4) ZeroShot Detection Examples after Unsupervised Pretraining on in-domain Examples followed by Adaptation [AUCROC : 0.8614748819235724]**

Correct prediction example which other strategies miss

#hayali will #asylanten nach spanien abschieben, wo über 20% arbeitslose sind und eine riesige verschuldung! mein tipp: Grenzen sichern!

### **Examples where all models are wrong**

Merkel sagte gestern bei einer Rede es würden ja keine Massen kommen sondern nur einzelne... Die Frau muß völlig durchgeknallt sein. Wenn wir nichts tun geht das hier nicht mehr lange gut. Am besten wäre wir machen uns alle auf den Weg nach Berlin!

schmeisst die krimminelen ausländer sofort raus,die haben hier in deutschland nichts verloren

## **Example Italian Social Media Posts**

### **1 ) ZeroShot Detection Examples [AUCROC 0.6087089978583474]**

False Positive example (other strategies predicted correctly)

Roma, lo sgombero di migranti arriva in Campidoglio. Gli attivisti: "Non ci fermiamo"  
@LaStampa <https://t.co/hWI9h8uDxO>

True Positive ( other strategies got this right )

@cicapingui @Piovegovernolad @ilPontormo #no c'è bisogno di fare figli X sconfiggere l'islam, basta non farlo entrare nel nostro paese

True Negative (other's incorrect)

@sten\_lomonaco @LuigiRuggerone L'Italia è ammazzata dalle tasse, ma vogliono farci credere che il nemico è il terrorismo. [ **Interesting example** ]

## **2 ) FewShot detection Examples after Supervised fine tuning on 256 in-domain examples [AUCROC 0.8351242683659832]**

True Negative Example (others false positive)

@senateurJGM @EnricoLetta Gentiloni, da Libia ove ls sembra contare meno proviene maggior numero migranti. Lei dov'è, a fare campagna per sì?

## **3) ZeroShot Detection Examples after Unsupervised Pretraining on in-domain Examples [ AUCROC 0.6628467055315324]**

**Example which this strategy predicted correctly while others failed**

#dallavostraparte Ha fatto bene il ristorante dare cena a soli italiani, sono i più bisognosi rispetto ai migranti che il Gov.PD da tutto.

## **4) FewShot Detection Examples after Unsupervised Pretraining on in-domain Examples followed by supervised pretraing on 256 in domain examples [AUCROC 0.8697210016164517]**

True negative (others incorrect )

#agorai parlano ancora di cazzate, dei clandestini tra cui ci sono solo il 10% di migranti x guerra. Gli altri sono voti x il PD. [ Got this right inspite of distractor keywords]

True Positive [ others incorrect ]

Quel figlio di Vanna marchi....se la pigliasse ne ano più profondo

### **Example where all strategy predicts incorrectly**

#papamilano2017 abbraccia gli #islamici, un abbraccio fatale che distrugge 2000anni di storia #cristiana. Sta aprendo le porte alla fine.

Accampamento rom improvvisato, 5 denunce per allaccio abusivo alla rete elettrica  
<https://t.co/ltSeolf42k>

VIDEO > Pazzesco!! Tra auto di lusso e ville i rom nullatenenti ricevono dal comune anche 245 euro a persona! [ Annotation Issue ? ]

Sarebbe diventato il presidente più amato di tutti i tempi se invece dei musulmani avesse messo al bando i risvoltini ai jeans #muslimban [ sarcasm ? ]

Lo dice Sartori: "Integrare gli islamici è un'illusione della sinistra. Il multiculturalismo non esiste [ very hard example ]

Fare una legge a Protezione TOTALE delle Donne non solo Islamiche quando denunciano le violenze domestiche così da smascherare il falso islam [ Zeroshot got it right while adapted models predicted incorrectly ]

## **SPANISH SOCIAL MEDIA POSTS**

### **1 ) ZeroShot Detection Examples [ AUCROC 0.6489227465047349 ]**

Correct detection Examples:

Se pinta los labios de color de negro porque la chupa de muerte. [ All strategies predict this correctly ]

## **2 ) FewShot detection Examples after Supervised fine tuning on 256 in-domain examples [ AUCROC- 0.744836510530612 ]**

Incorrect Prediction Examples

#ChiringuitodeMega eres medio subnormal lobo Carrasco. Pero sin el medio. Vete a dar por el culo al amorfo lelo missing

## **3) ZeroShot Detection Examples after Unsupervised Pretraining on in-domain Examples [ AUCROC: 0.7566750730555222]**

Example where only Strategy 3,4 predict correctly

Según los monguers podemitas y demás: Los musulmanes matan a cristianos, y los malos son los cristianos y les acusan de islamófobos.

## **4) FewShot Detection Examples after Unsupervised Pretraining on in-domain Examples followed by supervised pretraing on 256 in domain examples [AUCROC : 0.8348866877957681]**

Correct detection Example where other strategies fail:

Lo voy a decir lo más suave que puedo. Si dices que @CiudadanosCs es un partido fascista eres un ignorante, un ciego y un gilipollas. #EleccionesCataluna

## **Example where all strategy predict incorrectly**

Hsm no son nazis, solo ayudan, para flipar la hipoc... — Chau, puedo saber porque me sigues?? Creo que mi cu... <https://t.co/fN9bUINZmr>

@As\_TomasRoncero Pero que mierda estas hablando subnormal de los cojones  
MAMAAAAAAA GORDOOOOOOOO.

Me cago en el puto feminismo de mierda menuda dictadura joder ojalá meteros una polla  
por el culo es que me cago en... <https://t.co/O9BeSRfIGN>

Solo puede pasar en el Régimen d España @PPopular -Borbón invitar a un asesino  
franquista a cerebración d Democracia <https://t.co/2NRdDRQ0dq>

### **Some Fringe Social Media Platforms which promotes radical content**

Whisper, Telegram, Parler, Gab, Stormfront