

# Classification of twitter data from the 2012 Emilia-Romagna earthquake by machine learning: comparison of k-nearest neighbours, kernel support vector machine, and string kernel methods

Jeffrey Michael Franc<sup>\*§</sup>  
Pier Luigi Ingrassia<sup>§</sup>  
Ester Boniolo<sup>§</sup>  
Luca Carenzo<sup>§</sup>  
Francesco Della Corte<sup>§</sup>

<sup>\*</sup>Department of Emergency Medicine, The University of Alberta, Edmonton, Alberta, Canada

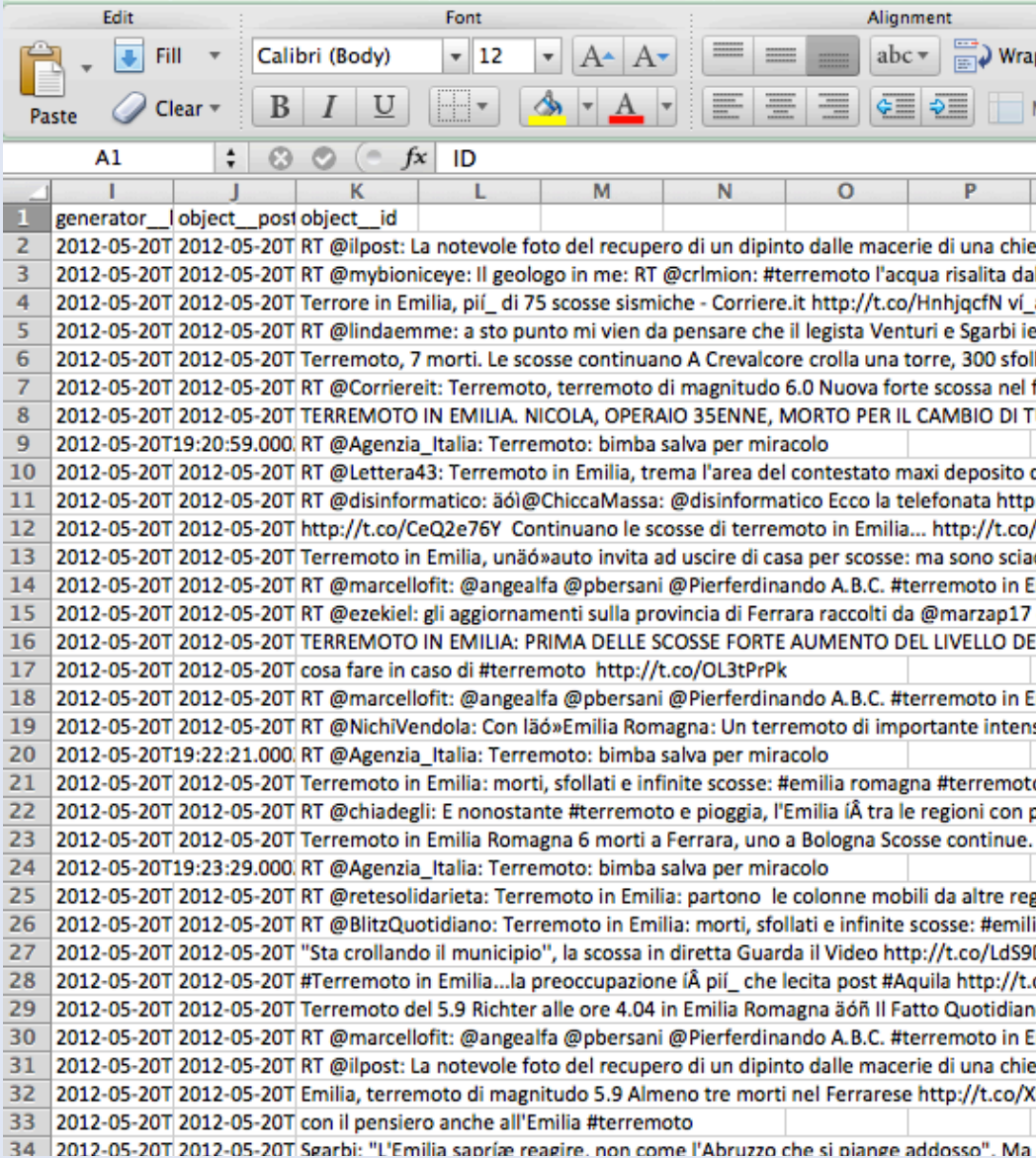
<sup>§</sup>Department of Translational Medicine, Università degli Studi del Piemonte Orientale "A. Avogadro", Novara, Italy.

MedStatStudio



# Introduction

- Twitter data may be useful for study of disasters
- The data is unstructured and high volume
- Human interpretation of the data is time consuming
- Machine learning methods are available



	I	J	K	L	M	N	O	P
1	generator_id	object_post	object_id					
2	2012-05-20T	2012-05-20T	RT @ilpost: La notevole foto del recupero di un dipinto dalle macerie di una chie					
3	2012-05-20T	2012-05-20T	RT @mybioniceye: Il geologo in me: RT @crlmion: #terremoto l'acqua risalita dal					
4	2012-05-20T	2012-05-20T	Terrore in Emilia, piú di 75 scosse sismiche - Corriere.it <a href="http://t.co/HnhjqcFN">http://t.co/HnhjqcFN</a> vi					
5	2012-05-20T	2012-05-20T	RT @lindaemme: a sto punto mi vien da pensare che il legista Venturi e Sgarbi ie					
6	2012-05-20T	2012-05-20T	Terremoto, 7 morti. Le scosse continuano A Crevalcore crolla una torre, 300 sfoll					
7	2012-05-20T	2012-05-20T	RT @Corriereit: Terremoto, terremoto di magnitudo 6.0 Nuova forte scossa nel f					
8	2012-05-20T	2012-05-20T	TERREMOTO IN EMILIA. NICOLA, OPERAIO 35ENNE, MORTO PER IL CAMBIO DI T					
9	2012-05-20T19:20:59.000		RT @Agenzia_Italia: Terremoto: bimba salva per miracolo					
10	2012-05-20T	2012-05-20T	RT @Lettera43: Terremoto in Emilia, trema l'area del contestato maxi deposito c					
11	2012-05-20T	2012-05-20T	RT @disinformatico: ãó@ChiccaMassa: @disinformatico Ecco la telefonata <a href="http://t.co/CeQ2e76Y">http://t.co/CeQ2e76Y</a>					
12	2012-05-20T	2012-05-20T	Continuano le scosse di terremoto in Emilia... <a href="http://t.co/">http://t.co/</a>					
13	2012-05-20T	2012-05-20T	Terremoto in Emilia, unãó»auto invita ad uscire di casa per scosse: ma sono scia					
14	2012-05-20T	2012-05-20T	RT @marcellofit: @angealfa @pbersani @Pierferdinando A.B.C. #terremoto in E					
15	2012-05-20T	2012-05-20T	RT @ezekiel: gli aggiornamenti sulla provincia di Ferrara raccolti da @marzap17					
16	2012-05-20T	2012-05-20T	TERREMOTO IN EMILIA: PRIMA DELLE SCOSSE FORTE AUMENTO DEL LIVELLO DE					
17	2012-05-20T	2012-05-20T	cosa fare in caso di #terremoto <a href="http://t.co/OL3tPrPk">http://t.co/OL3tPrPk</a>					
18	2012-05-20T	2012-05-20T	RT @marcellofit: @angealfa @pbersani @Pierferdinando A.B.C. #terremoto in E					
19	2012-05-20T	2012-05-20T	RT @NichiVendola: Con lâó»Emilia Romagna: Un terremoto di importante intens					
20	2012-05-20T19:22:21.000		RT @Agenzia_Italia: Terremoto: bimba salva per miracolo					
21	2012-05-20T	2012-05-20T	Terremoto in Emilia: morti, sfollati e infinite scosse: #emilia romagna #terremoto					
22	2012-05-20T	2012-05-20T	RT @chiadegli: E nonostante #terremoto e pioggia, l'Emilia iÃ tra le regioni con p					
23	2012-05-20T	2012-05-20T	Terremoto in Emilia Romagna 6 morti a Ferrara, uno a Bologna Scosse continue.					
24	2012-05-20T19:23:29.000		RT @Agenzia_Italia: Terremoto: bimba salva per miracolo					
25	2012-05-20T	2012-05-20T	RT @retesolidarieta: Terremoto in Emilia: partono le colonne mobili da altre reg					
26	2012-05-20T	2012-05-20T	RT @BlitzQuotidiano: Terremoto in Emilia: morti, sfollati e infinite scosse: #emili					
27	2012-05-20T	2012-05-20T	"Sta crollando il municipio", la scossa in diretta Guarda il Video <a href="http://t.co/LdS9f">http://t.co/LdS9f</a>					
28	2012-05-20T	2012-05-20T	#Terremoto in Emilia...la preoccupazione iÃ piú che lecita post #Aquila <a href="http://t.co/">http://t.co/</a>					
29	2012-05-20T	2012-05-20T	Terremoto del 5.9 Richter alle ore 4.04 in Emilia Romagna ãó Il Fatto Quotidian					
30	2012-05-20T	2012-05-20T	RT @marcellofit: @angealfa @pbersani @Pierferdinando A.B.C. #terremoto in E					
31	2012-05-20T	2012-05-20T	RT @ilpost: La notevole foto del recupero di un dipinto dalle macerie di una chie					
32	2012-05-20T	2012-05-20T	Emilia, terremoto di magnitudo 5.9 Almeno tre morti nel Ferrarese <a href="http://t.co/X">http://t.co/X</a>					
33	2012-05-20T	2012-05-20T	con il pensiero anche all'Emilia #terremoto					
34	2012-05-20T	2012-05-20T	Sgarbi: "L'Emilia sa piú reagire: non come l'Abruzzo che si piange addosso". Ma					

# Hypothesis and Testing

$H_0$ : There is no difference in accuracy of classification between:

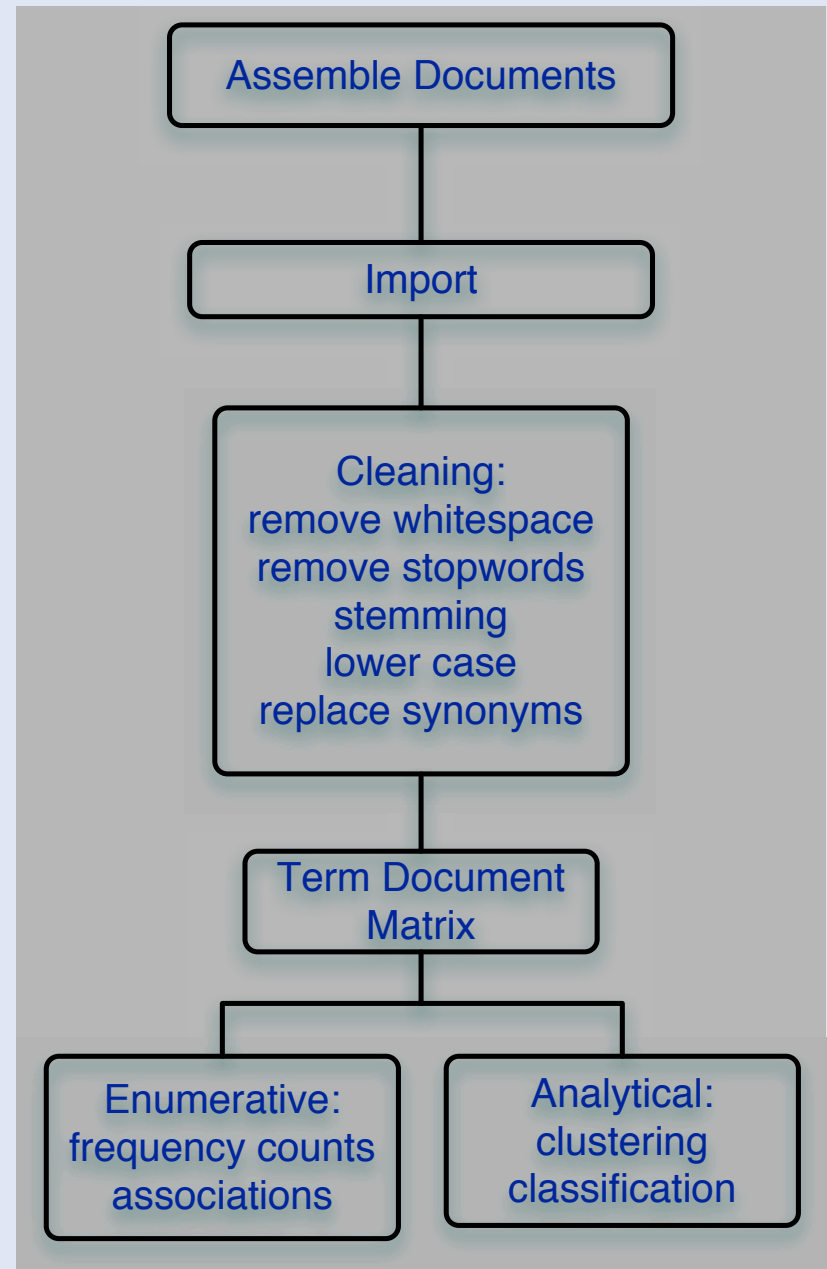
1. k-nearest neighbors (KNN)
2. Kernel support vector (KSVM)
3. String kernel support vector machine (SKSVM)

$H_a$ : There is significant difference between the three methods

*Tested by means of three sample proportion test*

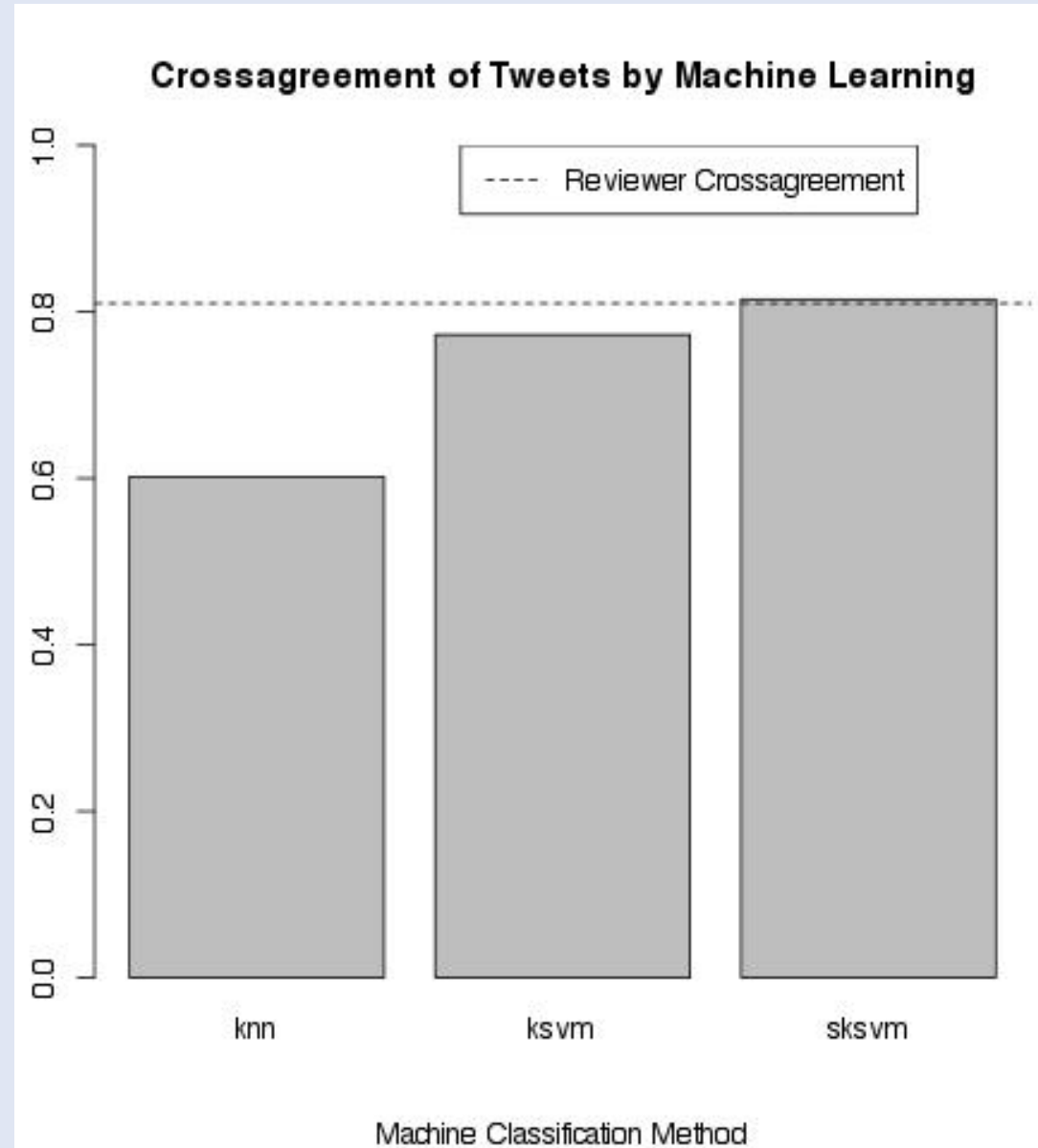
# Methods

- Tweets from first 24 hours after Emilia-Romagna earthquake (2012)
- Trained reviewers classified each tweet into one of five categories:
  1. Seeking information
  2. Giving general information
  3. Opinion, sentiment, or emotion
  4. Media or organizational information
  5. Links to other information



# Results

- 14,190 Tweets
- Cross agreement between two initial reviewers=82%.
- Cross agreement with machine learning:
  1. KNN: 62%
  2. KSMV: 78%
  3. SKSVM: 81%



# Conclusions

- String kernel support vector machine methods are most accurate machine learning ( $p < 0.0001$ )
- Nearly as accurate as human raters
- Show promise as methods of processing large amounts of twitter data
- Processor time is significant: 14 hours on 2.8ghz Intel i7/OSX/8gb ram.

# Appendix: Machine Learning Methods

Language: R Language for Computing

Packages:

1. tm (text mining)
2. class (function for classification)
3. kernlab (kernel based machine learning)
4. irr (inter-rater reliability)

Hardware: 2.8ghz Macintosh, 2-core Intel i7, 8gb ram

*All software freely available (GPL)*

*Source code for analysis available from  
the author*