Deep learning methods to detect radical online content

#TASMConf | SWANSEA UNIVERSITY | 26.06.19

Presented by

Dr. Berta Biescas

INSIKT
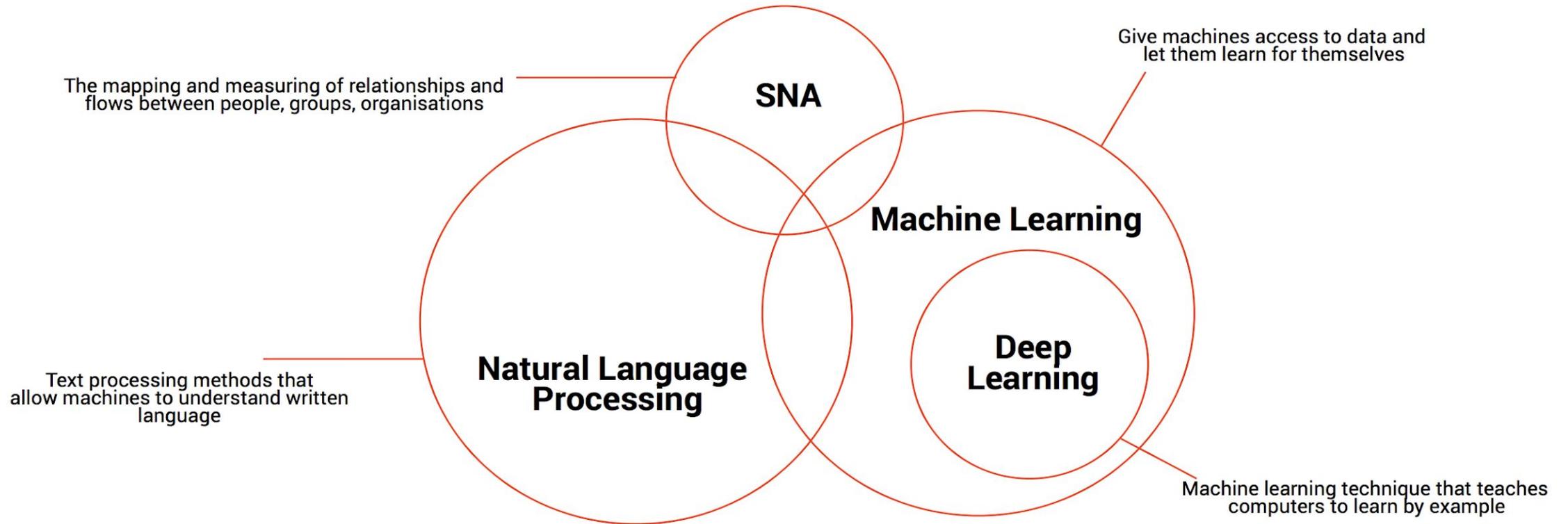INTELLIGENCE

insiktintelligence.com

INTRODUCTION: INVISO

Novel Social Data Mining Platform to Detect and Defeat Violent Online Radicalisation

INVISO

• Provides a novel solution for LEA analysts to detect terrorist propaganda on social media

• Identifies radical content, suspicious messages and covert radicalisation processes with the help of sophisticated text mining algorithms.

• Relies on deep learning to develop automatically new models which can be used to detect other criminal activity as well.

insiktintelligence.com

Scientific Pillars

SNA

Machine Learning

Natural Language Processing

Deep Learning

The mapping and measuring of relationships and flows between people, groups, organisations

Give machines access to data and let them learn for themselves

Text processing methods that allow machines to understand written language

Machine learning technique that teaches computers to learn by example

Analytical Methodologies

NATURAL LANGUAGE PROCESSING (NLP)

❏ Topic classification using deep learning domain

❏ Sentiment analysis using lexicon methods

❏ Key ideas based on pos tagging and linguistic theory

❏ Hashtags recognition

❏ Entities extraction

❏ Personality insights extraction.

❏ Concepts extraction using pos tagging and linguistic theory.

❏ Expertise in English, Arabic and Spanish languages.

❏ Deep Learning Models

❏ Transfer Learning

Analytical Methodologies

MACHINE LEARNING

❏ Text classification using supervised models, deep

learning and lexicon models.

❏ Word mapping design.

SOCIAL NETWORK ANALYSIS (SNA)

❏ Networks of concepts

❏ Analysis of text similarities.

❏ Detection of influencers.

❏ Analysis of clusters.

❏ Analysis of mentions,

retweets, replies, etc.
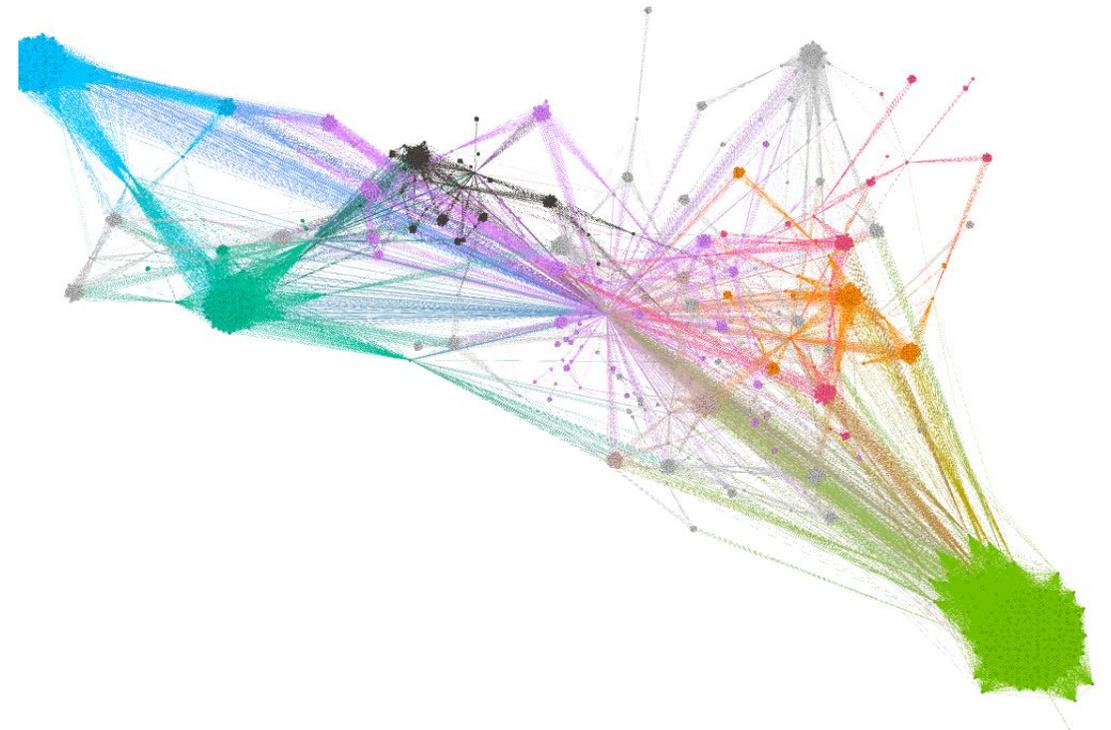
HOW INVISO HELPS

- INSIKT is a novel radicalisation detection technology, based on live semantic analysis of big data social media streams.

- To detect radicalisation activities very early (as they are taking place) and limit spread of radical content.

- To detect when terrorist or other radical groups are recruiting or releasing their propaganda.

- To detect suspicious user behaviour, identify hubs of radicalisation and intervene.
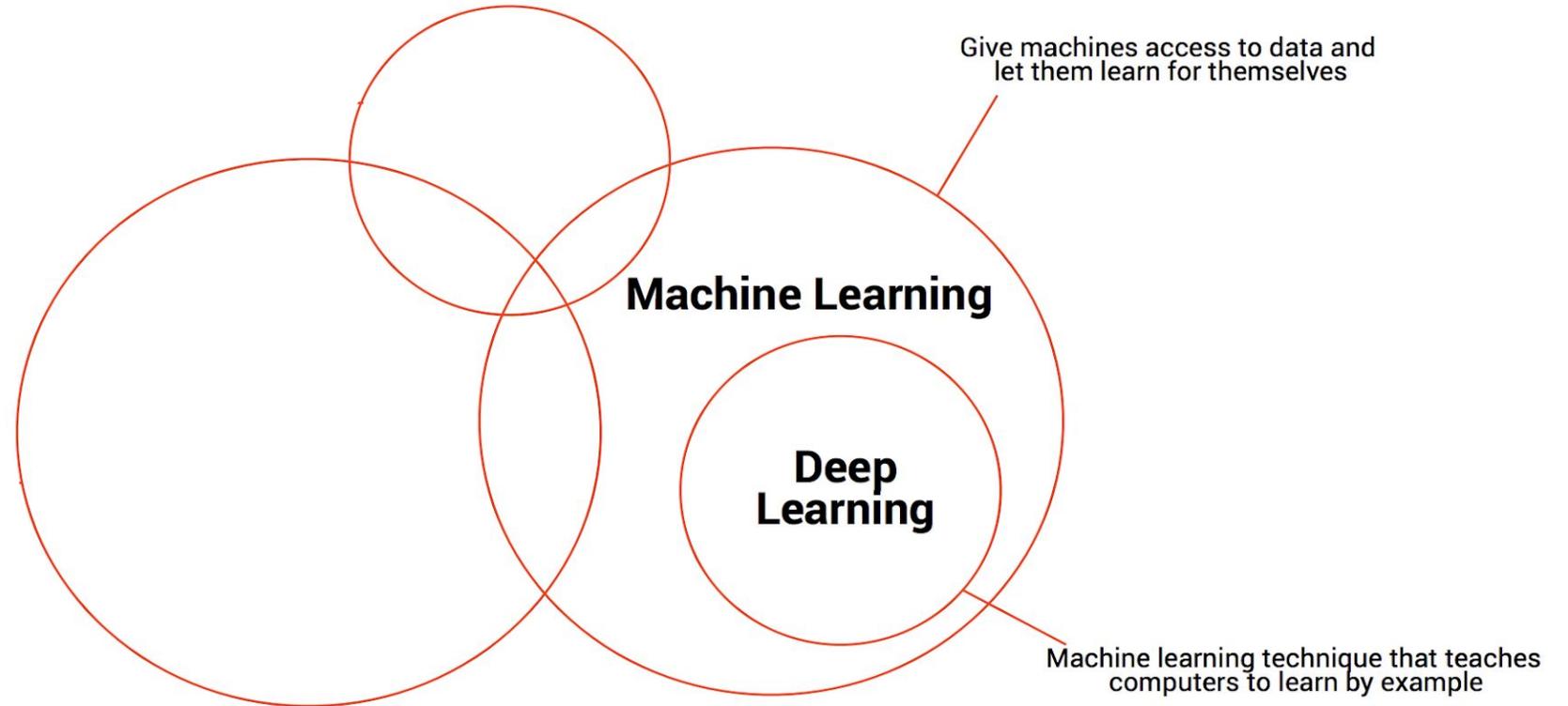

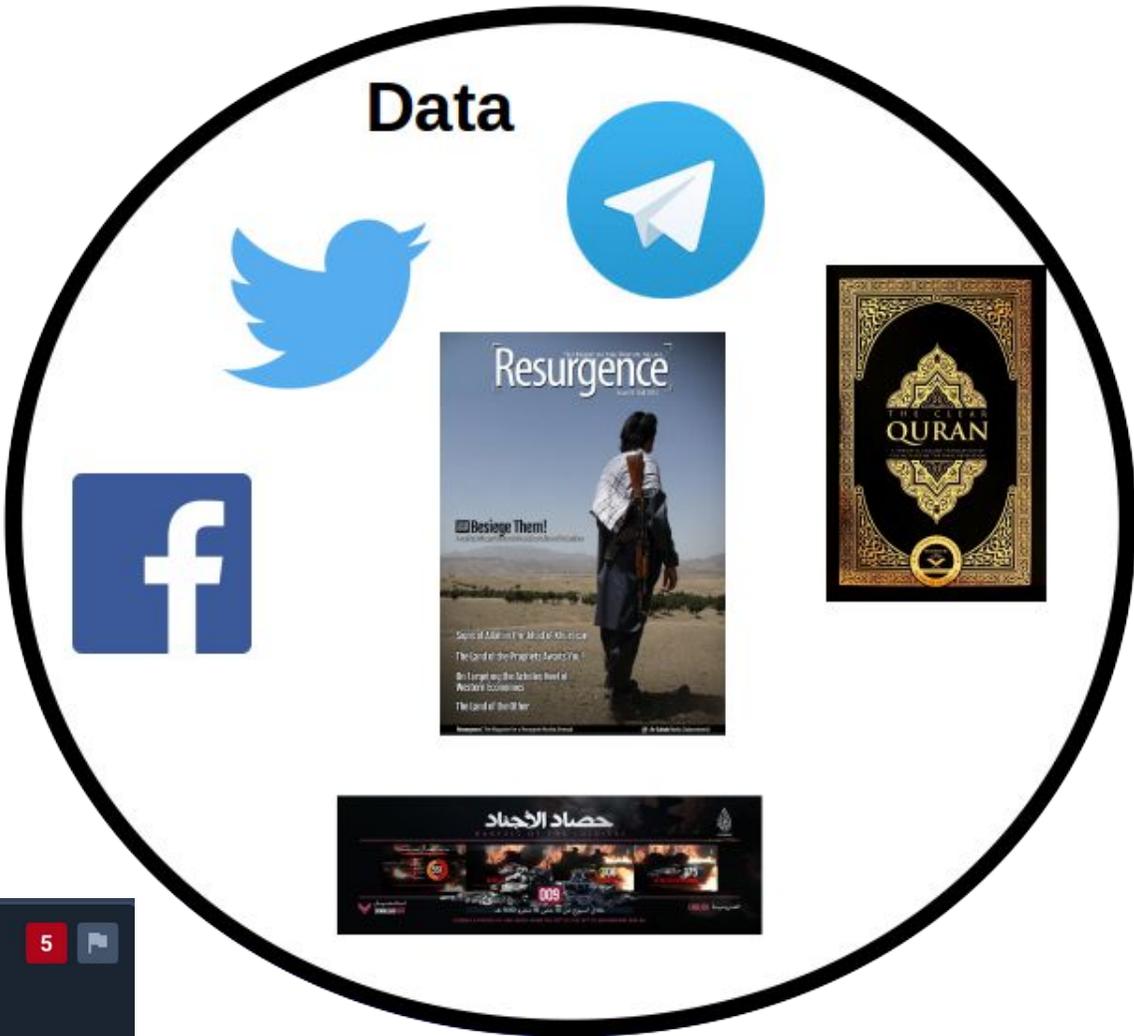**Early Detection of Radicalisation and Propaganda Content**

Detection of users exposed to radicalisation processes

insiktintelligence.com

Deep learning methods to detect radical content online

Give machines access to data and let them learn for themselves

Machine Learning

Deep Learning

Machine learning technique that teaches computers to learn by example

insiktintelligence.com

DEEP LEARNING

Data

Aaban Ismere @AabanISJih • 8 h

سُئِل النبله عليه وسلم أي العمل أحب إلى اللهلعمل أحب إلى الله ق قا على وقثم ثم براللع

SEE TRANSLATION

24,459 likes   37 retweets

$(23, 435,325,12,567, 54, 98, 90,...)$

model dimension

number of words

insiktintelligence.com
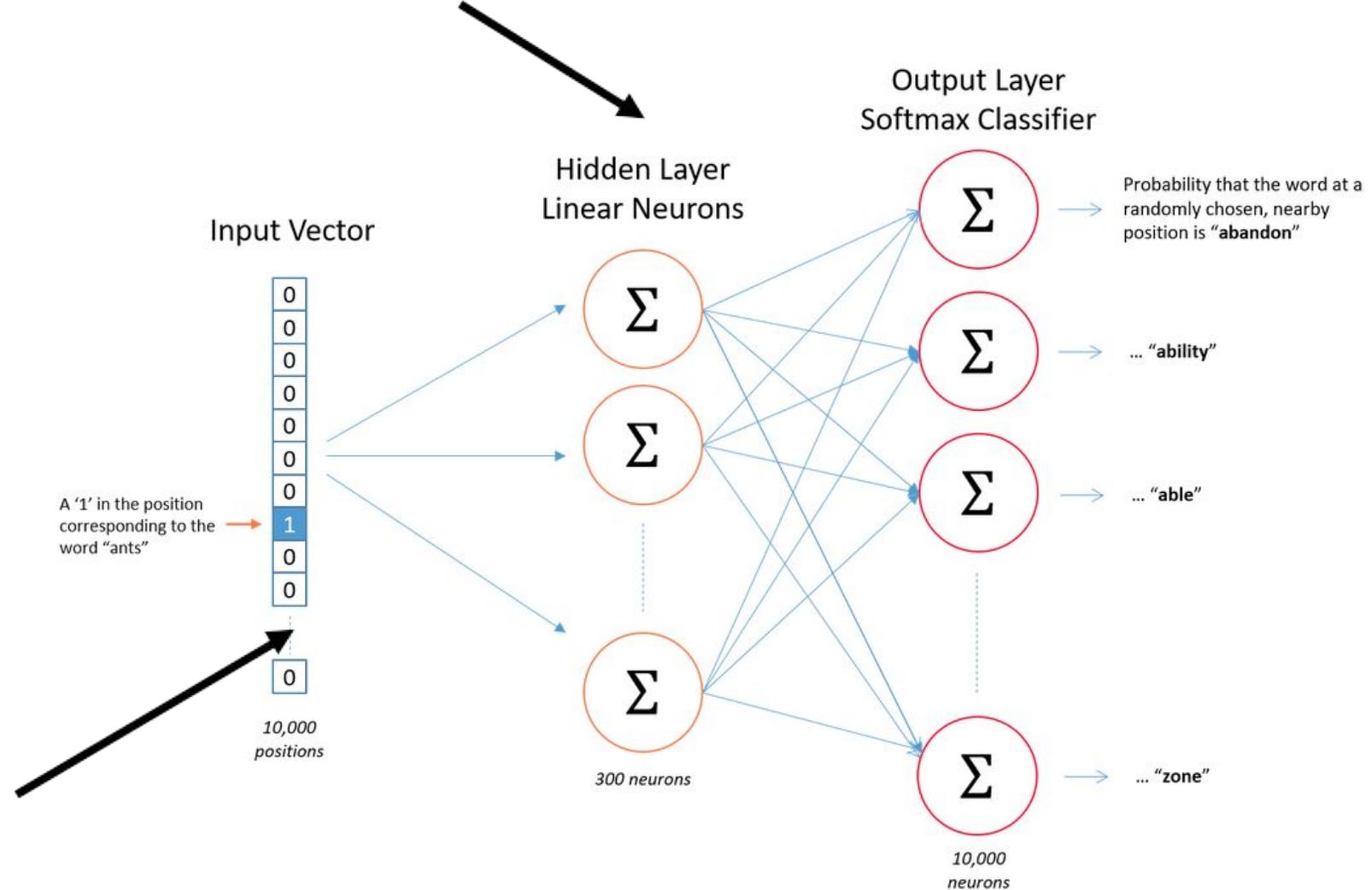
MACHINE LEARNING

TEXT CLASSIFIER

Supervised learning - Annotated Corpus

Non suspicious - L0

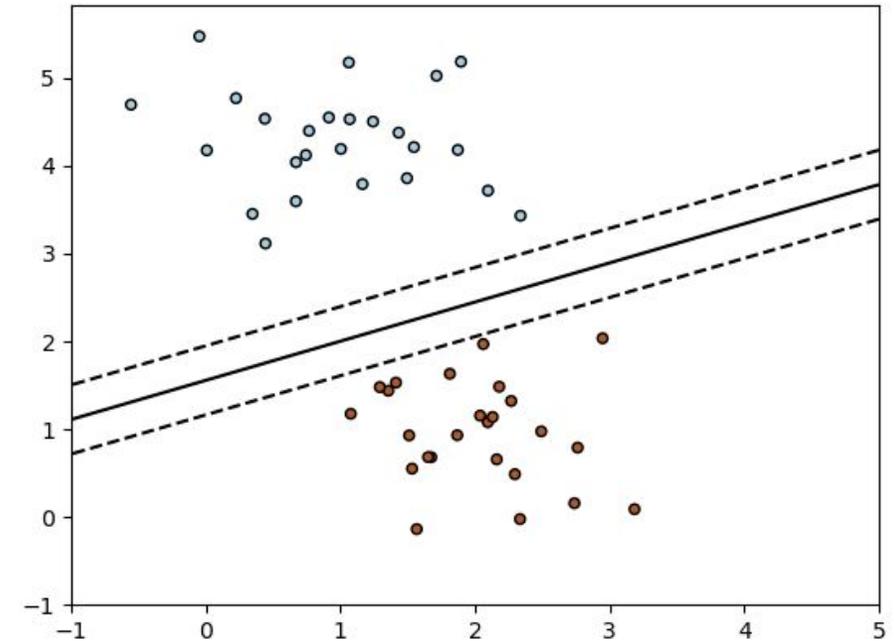Suspicious support for radical actions, recruitment, fundings - L1

Follow extremism ideas - L1

Joy for terrorism attacks/support radical people - L1

Call to action/support of terrorism attacks - L1

SGD CLASSIFIER

muslimah-UK @muslimahUK1 · Jun 11
their time is coming and then they will have to repent inshaAllah
💬 2   🔁 3   ♡ 15   ✉

L1

E ISLAM @eislaminfo · 5h
"Verily those who commit kufr, who forbid (others) from Allah's path (Islaam) and
then die as **Kuffaar**, Allah shall never forgive them".
Surah Muhammad, verse 34
💬   🔁   ♡   ✉

L1

Abu Bakr Bilyamin| @Serddieq · Jul 5
Replying to @Layorab @BITO_WALE @ICalvaries
Both Jews and Christians disbelieved in Islam, and therefore they are
disbelievers (**Kuffaar**).
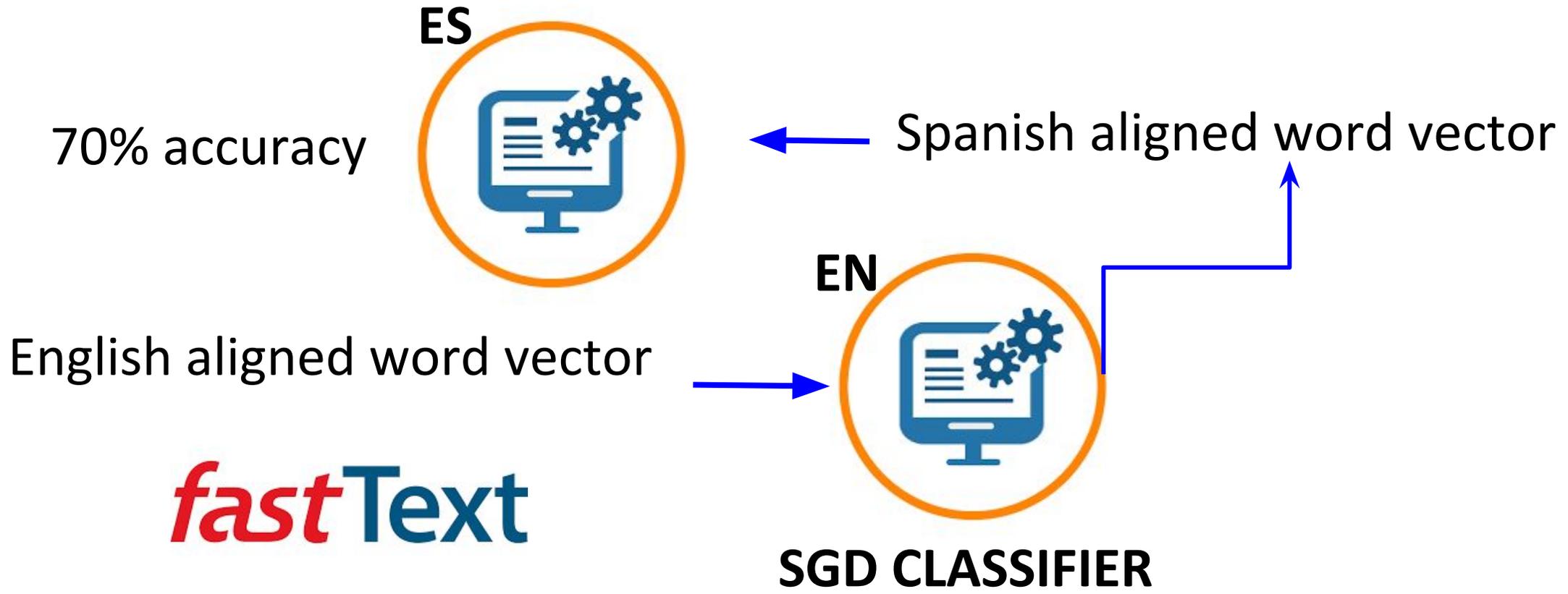💬 1   🔁   ♡   ✉

L1

insiktintelligence.com

Word vector	Accuracy English classifier	Accuracy Arabic classifier
generic	74.6 +- 2.2	77.7 +- 1
Jihadist domain	75.3+-1.2	79.6+-0.7

Transfer Learning + Aligned word vectors+ SGD CLASSIFIER

ES

70% accuracy ← Spanish aligned word vector

English aligned word vector →

EN

fastText

SGD CLASSIFIER

USER INSIGHTS

Threat Level Filter

124 New posts

Feed related to user: @AabanISJih

Twitter

@AabanISJih

CONTENT   USERS   LOCATIONS

Content related to user: @AabanISJih

Aaban Ismere @AabanISJih • 8 h

#IS is degrading and destroying America and its allies #Baghdad. 1000+ Islamic state fighters surrounding Baghdad. #calamitywillbefallus

1,364 likes   56 retweets

Aaban Ismere

Link > Twitter Account

Key ideas   POSTS

Bomb attacks   320,304

Weapons and explosives   120,987

Government   90,605

Recruiting   84,392

Terror   56,039

Followers   Following   Posts
3,456   254   1,234

Top key idea
Bomb attacks

Top topic
Recruiting

Top entity
Government

THREAT SCORE
3,7

Twitter

Aaban Ismere @AabanISJih • 12 h

سُئل النبي صلى الله عليه وسلم أي العمل أحب إلى اللهعمل أحب إلى الله ق
قا على وق قال ثم قال برالعمل أحب إلى الله قوالدين قال ثم قال أي الجهاد فيالال

SEE TRANSLATION

Jihadists London 293349 posts analysed

THREAT LEVEL 4
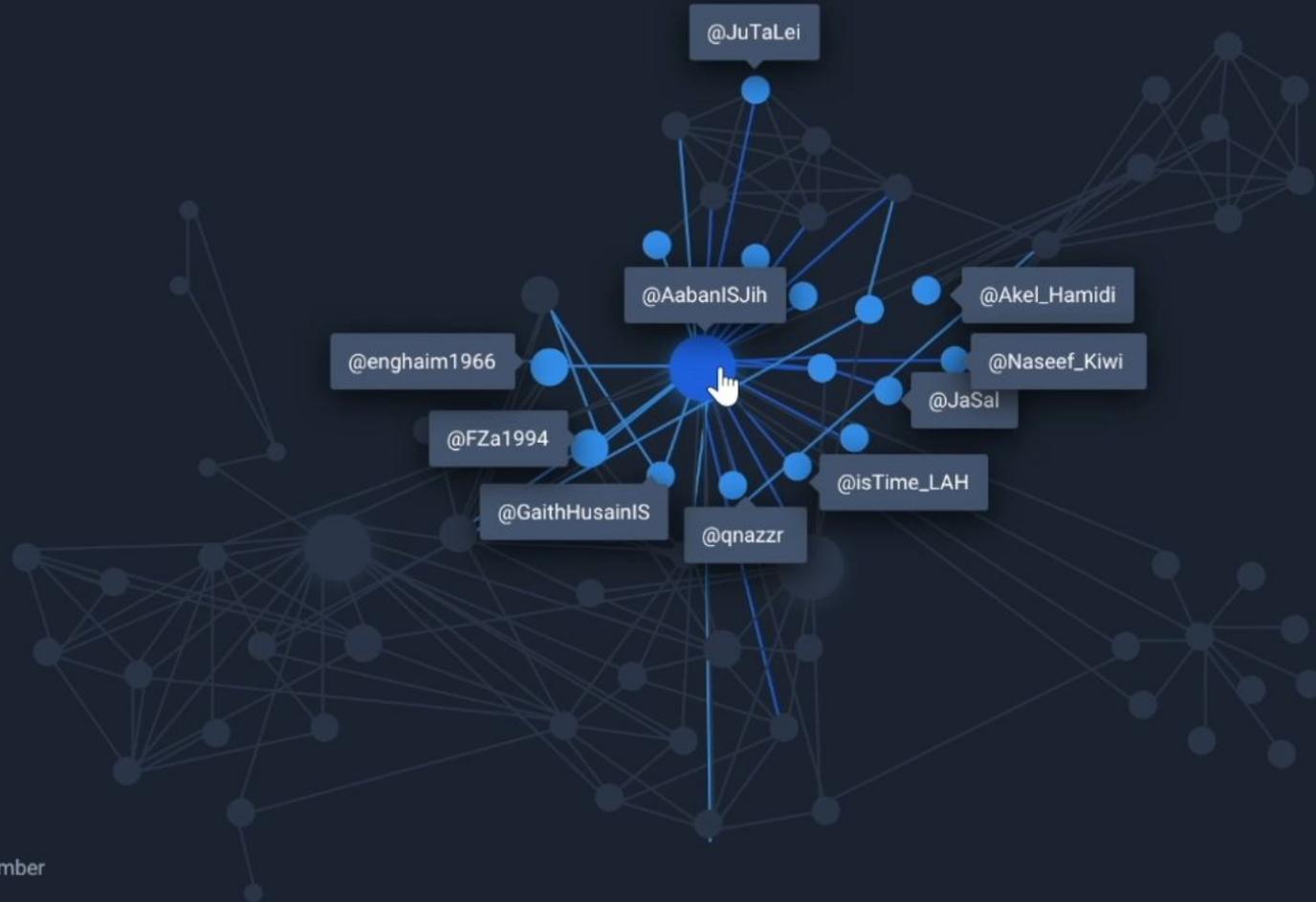
3/JAN/2019 13:00 h - 16/MAY/2019 17:37 h   KEYWORDS   FEED

Threat Level Evolution

Posts

Last 15 days

Select the threat levels to show:

- 5 - Critical Threat Level
- 4 - Severe Threat Level
- 3 - Substantial Threat Level
- 2 - Moderate Threat Level
- 1 - Low Threat Level

3,000

2,500

2,000

1,500

1,000

500

0

01 May 02 May 03 May 04 May 05 May 06 May 07 May 08 May 09 May 10 May 11 May 12 May 13 May 14 May 15 May 16 May

1: 2,284
2: 140
3: 1
4: 0
5: 1

Top Users	POSTS
1. Ultrascanhumint	301
2. Wedata_develope	285
3. NewsAt20	273
4. SWedata	204
5. Mitanurmousumi	203

Top Key Ideas	POSTS
1. Boko haram	27953
2. Haram in islam	6920
3. Hiv came	6530
4. Nebola came	6530
5. Stardust crusaders	6405

Top Retweets

Twitter

Gᴇᴋʏᴜᴍᴇ 全快@phxnso · a month ago

ATTENTION ALL STARDUST CRUSADERS
https://t.co/B99FB5MxaF

0 likes   3998 retweets

Twitter

ash 🍊 @ashloveslove · a month ago
Jotaro after star dust crusaders

THANK YOU

berta@insiktintelligence.com