

ANTISPAM & MACHINE LEARNING AT LOVOO

Juan De Dios Santos Rivera
Software Engineer Big Data
LOVOO

Berlin Buzzwords 2017
June 12, 2017

AGENDA

SPAM AT LOVOO

ANTISPAM

INSIGHTS

ABOUT LOVOO



Juan De Dios Santos Rivera
Software Engineer Big Data
LOVOO

Berlin Buzzwords 2017
June 12, 2017

WHAT IS SPAM?

UNAUTHORIZED

UNSOLICITED &
DISRUPTING

ADVERTISEMENT

GOAL IS TO GENERATE
revenue/views/any sort of profit



WHY DOES SPAM SUCK?

Due to the nature of our service, many spammers disguise themselves as **real users**. They use **genuine images and descriptions** in their profiles to convince our users into believing they are **real**.



NEGATIVE EFFECTS

DAMAGE THE IMAGE OF THE PRODUCT

DISTRACT OUR CUSTOMER SUPPORT

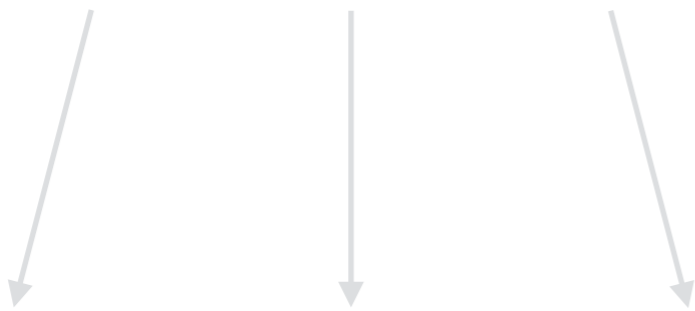
MAKE OUR USERS SAD 🙄
SO, THEY QUIT THE PLATFORM



HOW DOES A SPAMMER* BEHAVE

USE PICTURE OF ATTRACTIVE PEOPLE

(artists, Instagram, ...)



Spam ad



Blackmail



Prostitution

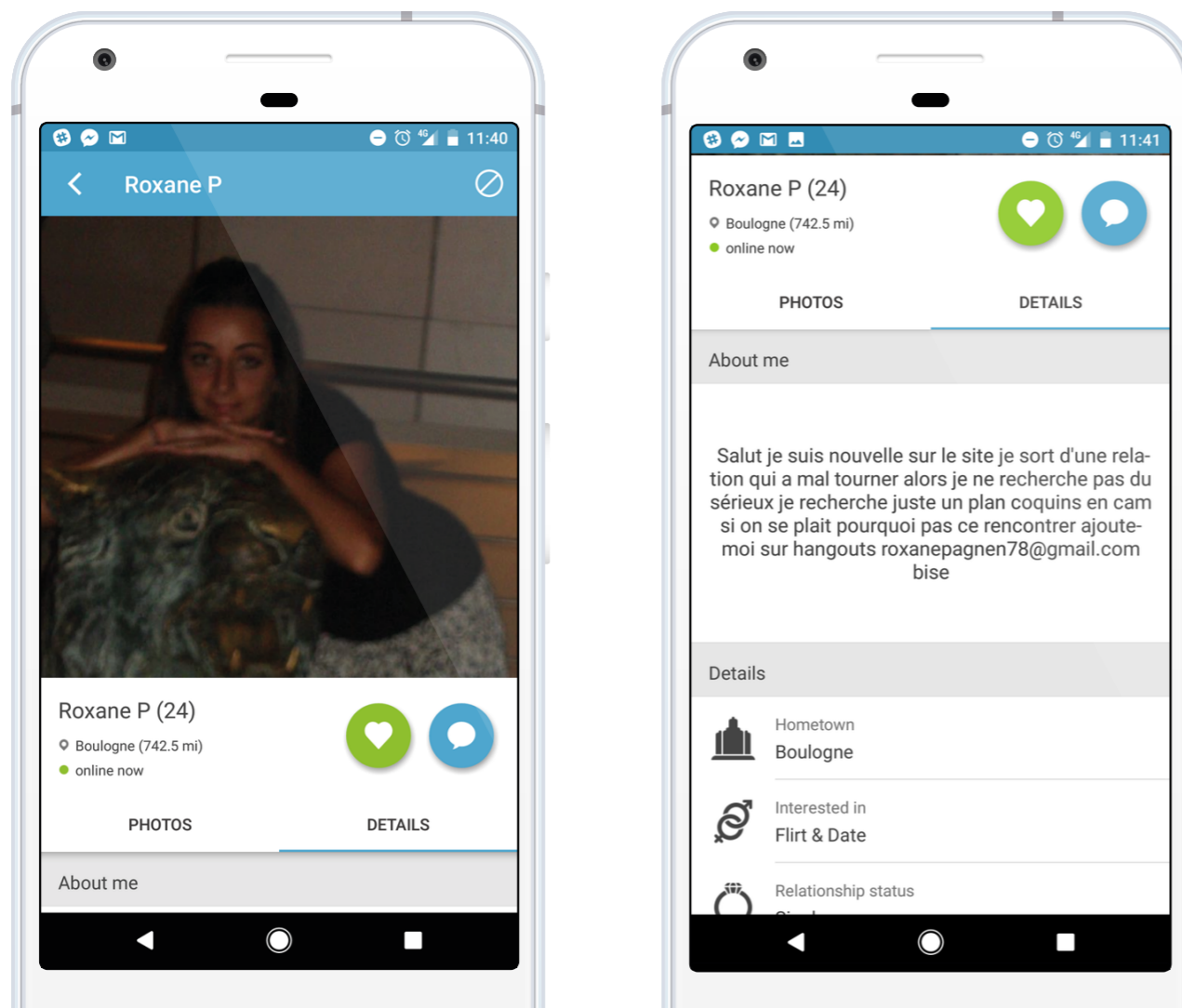
WRITE AN ADVERTISEMENT
IN THEIR STATUS



Vote!

*bots

SPAMMER PROFILES



ANTISPAM

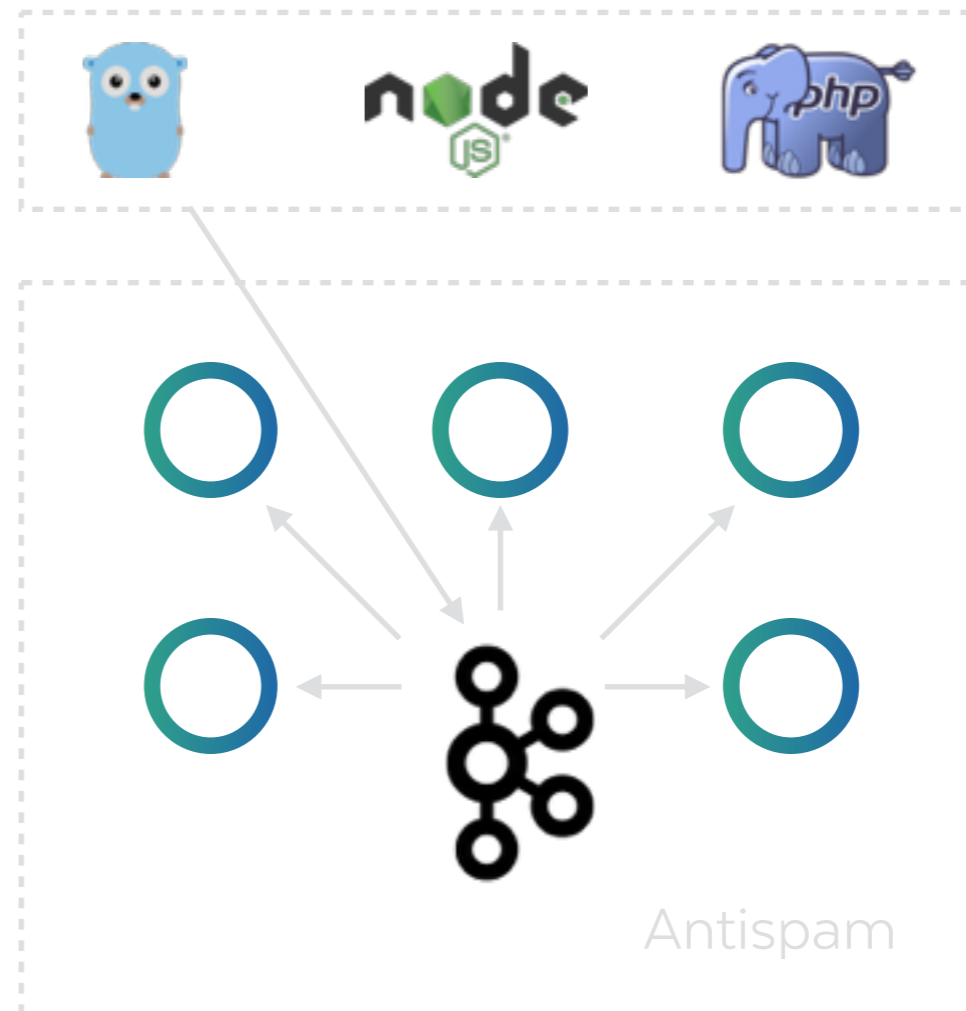


ANTISPAM

EVENT-BASED SYSTEM

IT COMMUNICATES
THROUGH **KAFKA** TOPICS

REAL-TIME PROCESSING



PILLARS OF ANTISPAM

NON MACHINE LEARNING BASED

Rule Based System
Community



MACHINE LEARNING BASED

Activity Based Model
Report Clustering
Entity Reputation



ACTIVITY BASED MODEL

SUPERVISED LEARNING APPROACH

RANDOM FOREST OF 800 TREES

CLASSIFIES A USER INTO A **SPAMMER** OR **HAMMER**

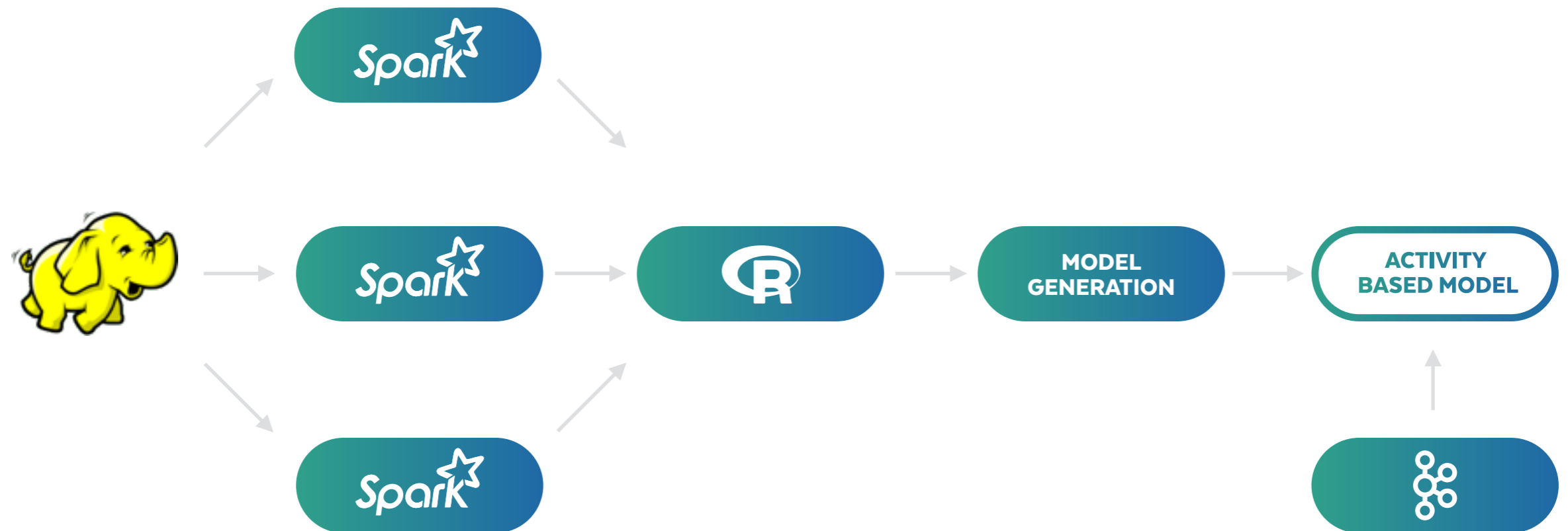
MODEL USES OVER 30 FEATURES RELATED TO THE HOW THE USER BEHAVES

SOME OF THE VALUES OF THESE FEATURES ARE THE ABSOLUTE NUMBER OF EXECUTION OF A CERTAIN EVENT, AND OTHERS ARE THE PROPORTION OF AN EVENT WITH RESPECT TO OTHERS

EXAMPLE OF FEATURES

userid	messages_blocked	was_blocked	today_logged_in	...
A	0.16	0.0	2	...
B	0.5	0.023	1	...
C
D

SYSTEM PIPELINE



REPORT CLUSTERING

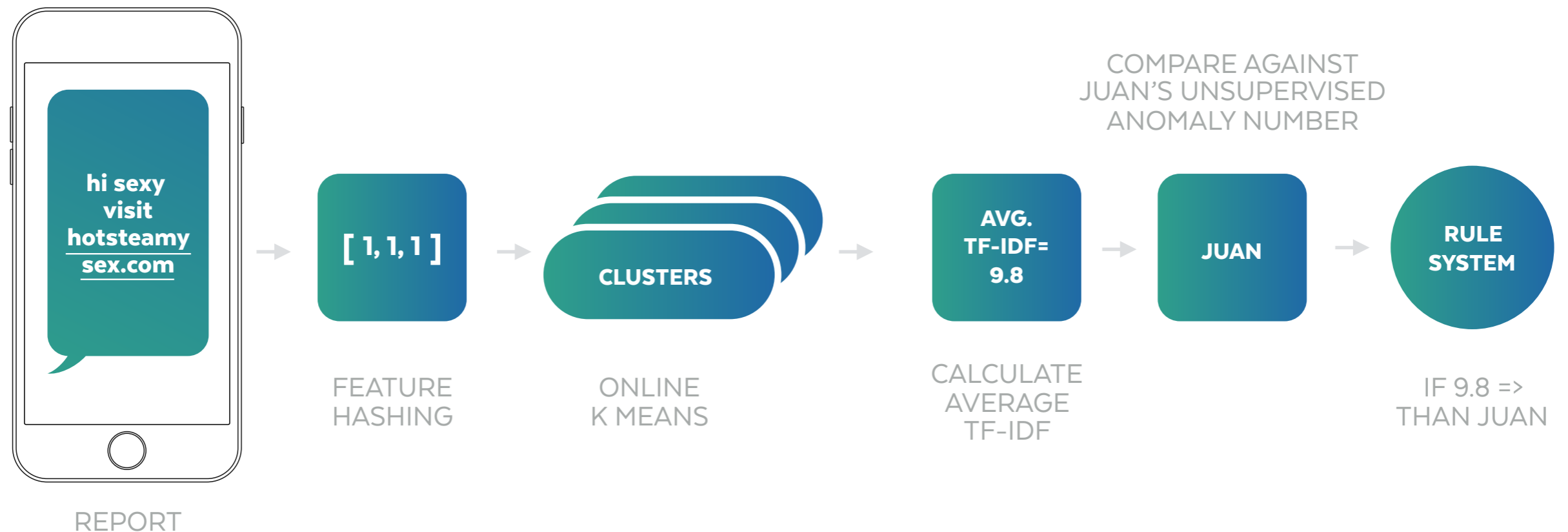
UNSUPERVISED LEARNING

ONLINE VERSION OF K-MEANS

ITS MAIN PURPOSE IS TO DETECT AND COMBAT SPAM WAVES

CREATE DYNAMIC RULES FOR THE RULE SYSTEM BASED ON CONTENT REPORTED BY THE USERS

REPORT CLUSTERING



ENTITY REPUTATION

SUPERVISED LEARNING

RANDOM FOREST

CLASSIFIES A USER INTO EITHER A
SPAMMER OR HAMMER

ENTITIES

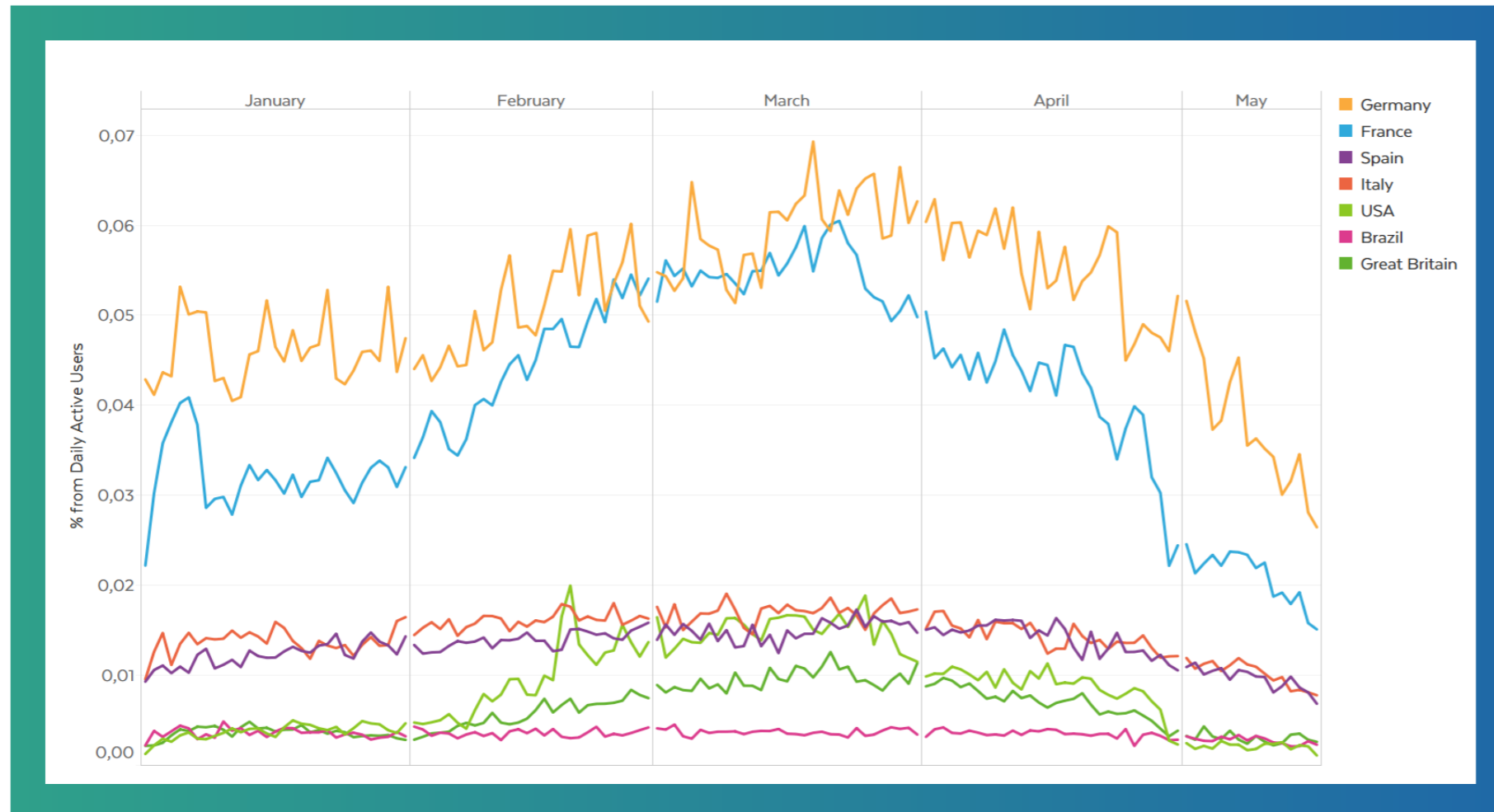
ENTITY = RESOURCES OR SYSTEMS A USER USES
TO COMMUNICATE TO THE PLATFORM



INSIGHTS



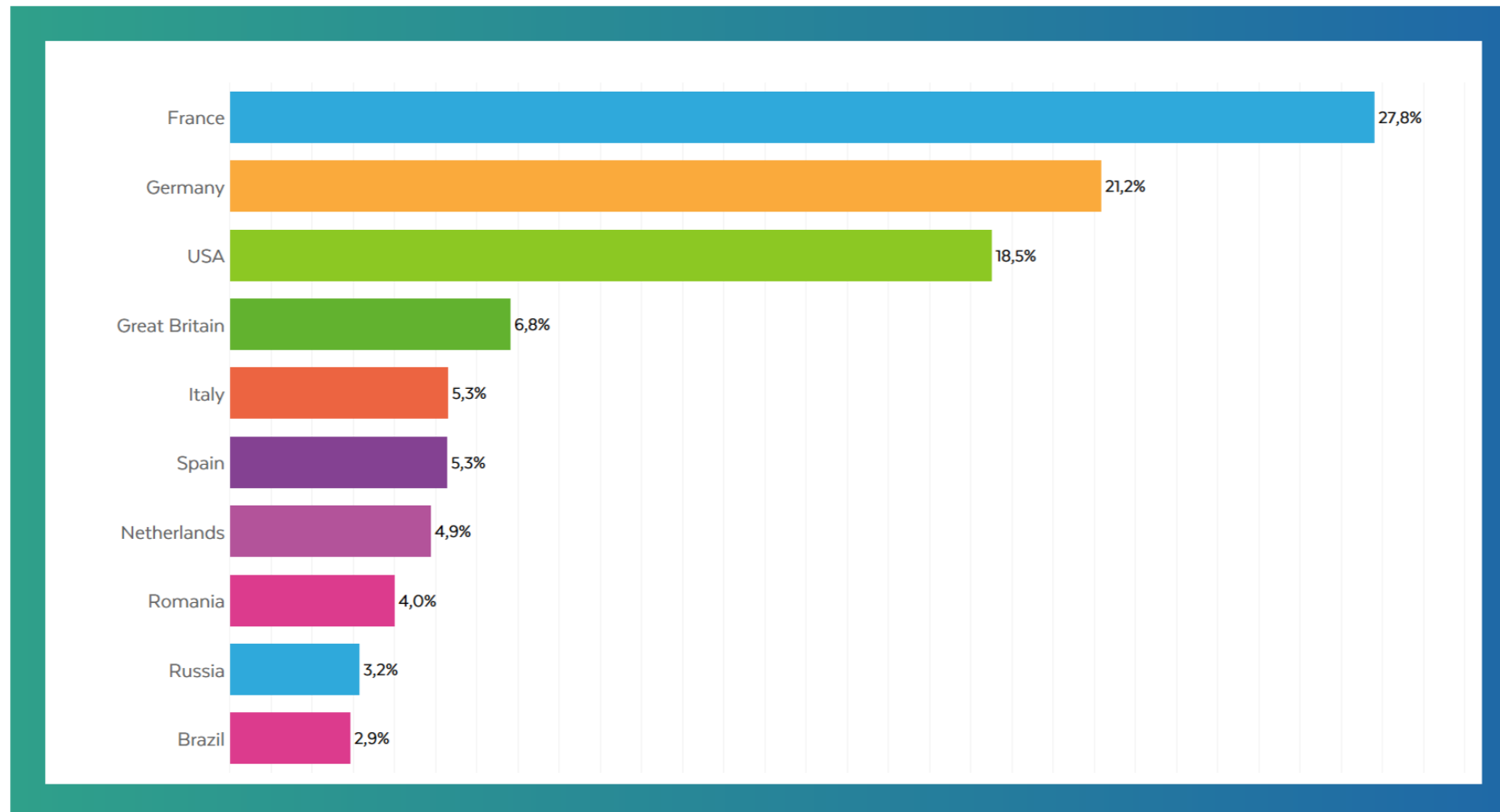
REPORTED USERS ACROSS MONTHS



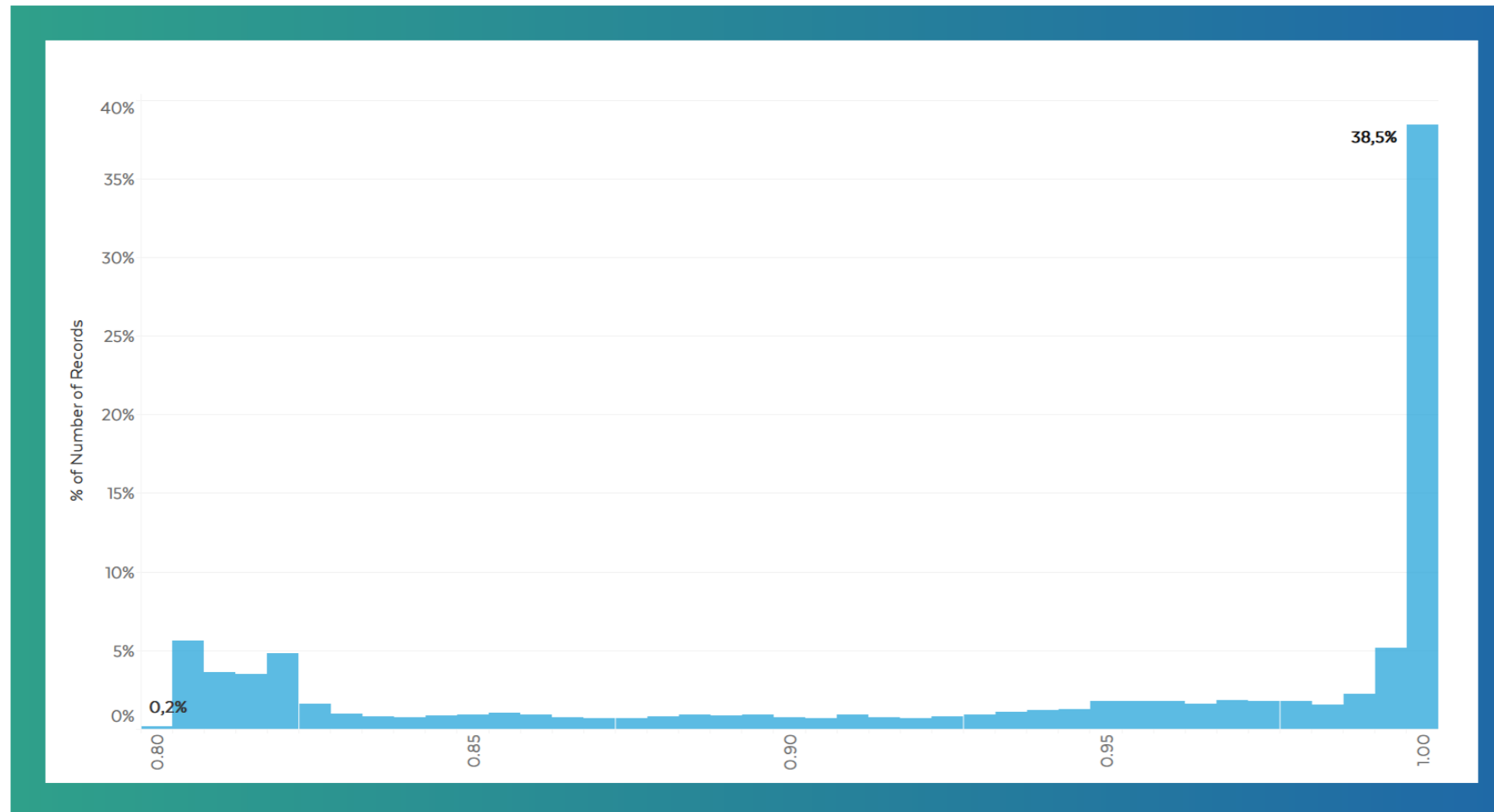
BLOCKED USERS



SPAMMERS BY COUNTRY



HISTOGRAM OF ACTIVITY BASED MODEL CONFIDENCE VALUES



THANKS

Say no to spam!

Twitter: @jdiossantos

LOVOO Engineering: @LOVOOEng