Anàlisi automàtica de tiquets de phishing

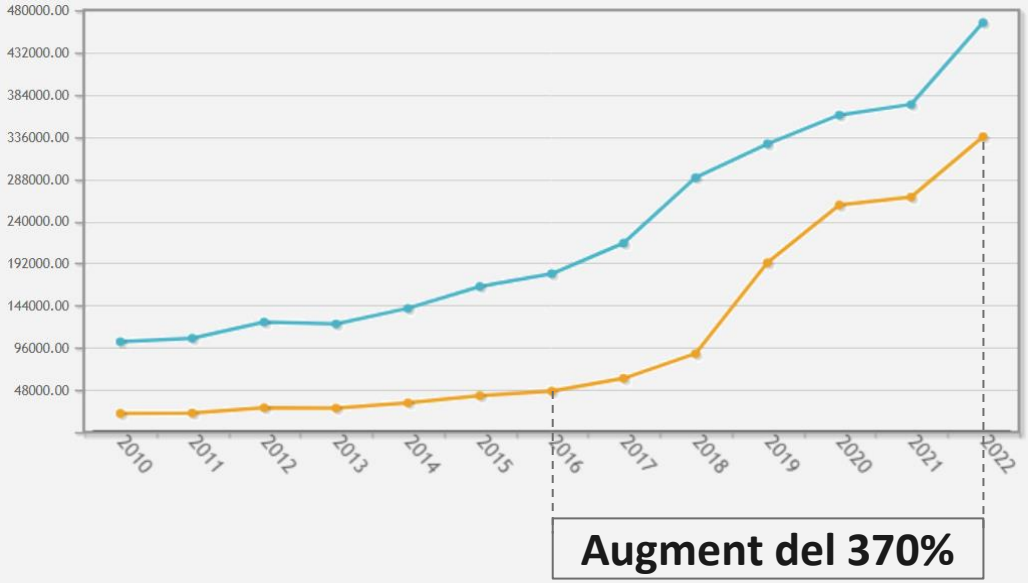
mitjançant el processament del llenguatge natural

Jaume Casals Vilaplana

01

Context i abast

Nombre d’estafes nacionals



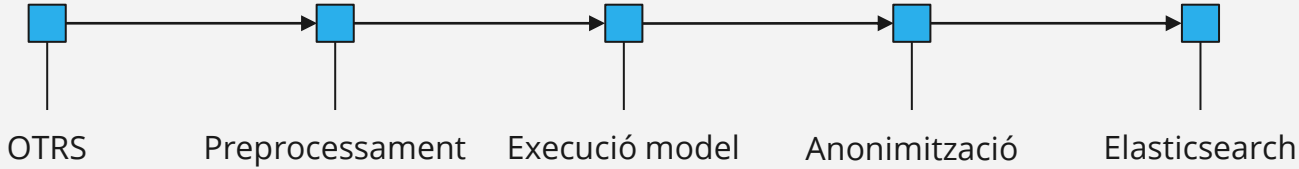
465.906 casos

Estafes

335.995 casos

Estafes informàtiques

Abast



Objectius

Estudi de l'estat de l'art

Eina descàrrega de tiquets

Preprocessament de les dades

Entrenament del model NLP

Anonimització de la sortida

Emmagatzematge del resultat

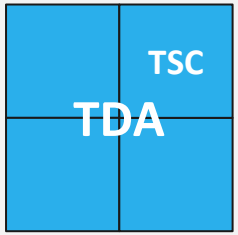
Implementar pipeline

Accés mitjançant API

Actors implicats

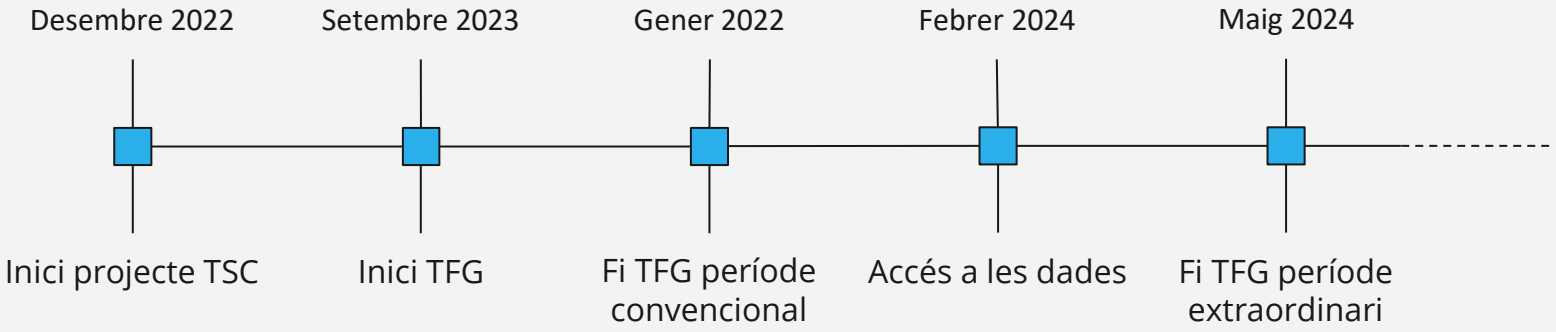
Agència

i2CAT



InLab FIB

Cronologia

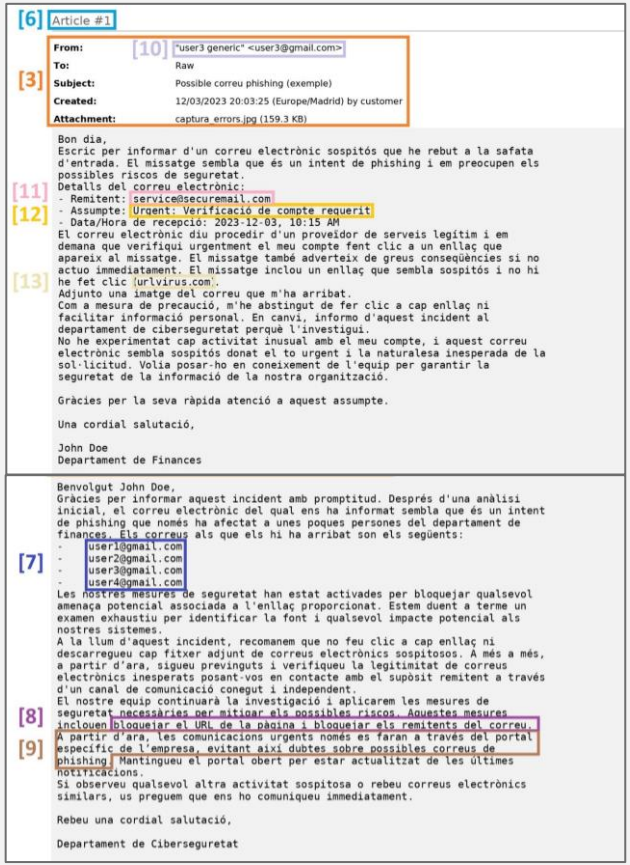

Desembre 2022
Setembre 2023
Gener 2022
Febrer 2024
Maig 2024
Inici projecte TSC
Inici TFG
Fi TFG període convencional
Accés a les dades
Fi TFG període extraordinari

02

Exploració teòrica

Anàlisi d’un tiquet

7. Usuaris afectats
8. Accions de mitigació
9. Accions de control
10. Mail de la víctima
11. Mail de l’atacant
12. Assumpte del correu
13. URL de l’incident



OTRS i Elasticsearch

Gestió de tiquets

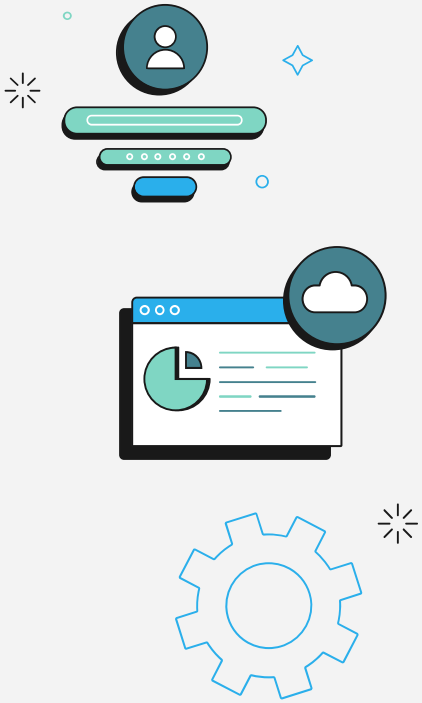
OTRS
Open Technology
Real Services

Znuny

Base de dades

elasticsearch

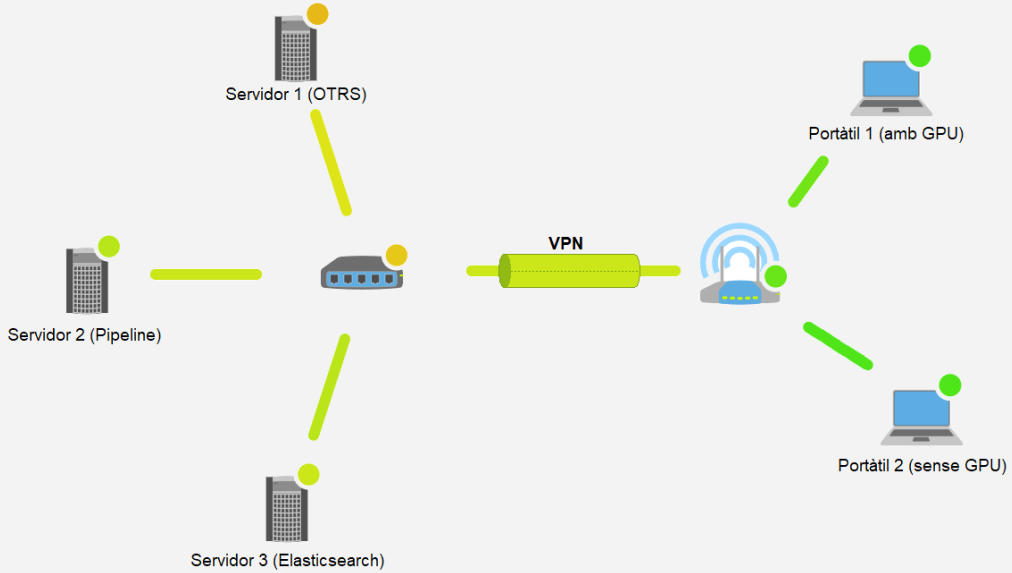
Logstash



03

Execució pipeline

Diagrama de la xarxa



Extracció de dades

Eina principal: PyOTRS

- Accés API
- Tiquet en format HTML

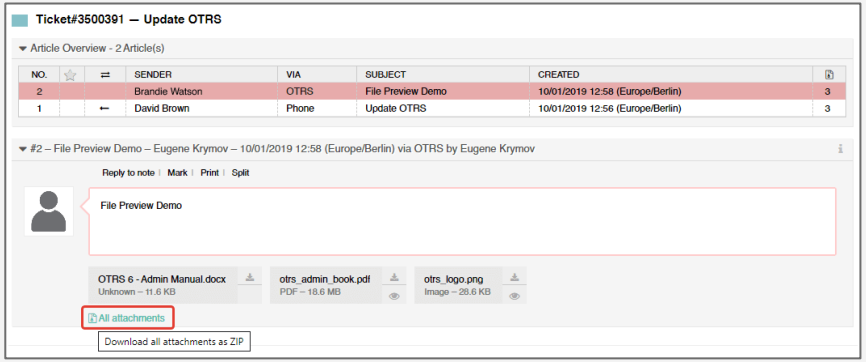
```
    </style>
  </head>
  <body>
    <table role="presentation" border="0" cellpadding="0" cellspacing="0" class="body">
      <tr>
        <td> </td>
        <td class="container">
          <div class="content">

            <!-- START CENTERED WHITE CONTAINER -->
            <span class="preheader">This is preheader text. Some clients will show this text as a
preview.</span>
            <table role="presentation" border="0" cellpadding="0" cellspacing="0" class="main">

              <!-- START MAIN CONTENT AREA -->
              <tr>
                <td class="wrapper">
                  <p>Bon dia,</p>
                  <p>Escric per informar d'un correu electrònic sospitós que he rebut a la safata
d'entrada. El missatge sembla que és un intent de phishing i em preocupen els possibles riscos de
seguretat.</p>
                  <p>Detalls del correu electrònic:</p>

<p>Remitent: service@securemail.com</p>

<p>Assumpte: Urgent: Verificació de compte requerit</p>

<p>Data/Hora de recepció: 2023-12-03, 10:15 AM</p>

<p>Adjunto una imatge del correu que m'ha arribat.</p>

<p>El correu electrònic diu procedir d'un proveïdor de serveis legítim i em demana que verifiqui
urgentment el meu compte fent clic a un enllaç que</p>

<p>apareix al missatge. El missatge també adverteix de greus conseqüències si no actuo immediatament.
El missatge inclou un enllaç que sembla sospitós i no hi he fet clic (urlvirus.com).</p>

<p>Com a mesura de precaució, m'he abstingut de fer clic a cap enllaç ni facilitar informació
```

Preprocessament

- Eliminar repeticions
- Detectar referències

Adjunts que es llegeixen

- Text
- Email
- PDF
- Word
- Word (antic)




```
Si observeu qualsevol altra activitat sospitosa o rebeu correus electrònics
similars, us preguem que ens ho comuniqueu immediatament.

Rebeu una cordial salutació,

Departament de Ciberseguretat

--
Super Support - Waterford Business Park
5201 Blue Lagoon Drive - 8th Floor & 9th Floor - Miami, 33126 USA
Email: hot@example.com - Web: [1]http://www.example.com/
--

12/03/2023 20:03 (Europe/Madrid) - user3 generic wrote: > Bon dia,
> Escric per informar d'un correu electrònic sospitós que he rebut a la safata
> d'entrada. El missatge sembla que és un intent de phishing i em preocupen els
> possibles riscos de seguretat.
> Detalls del correu electrònic:
> - Remitent: service@securemail.com
```

Eliminació de signatures

Flan-T5-base/FLOR

...

Si observeu qualsevol altra activitat sospitosa o rebeu correus electrònics similars, us preguem que ens ho comuniqueu immediatament.

Rebeu una cordial salutació,

Departament de Ciberseguretat

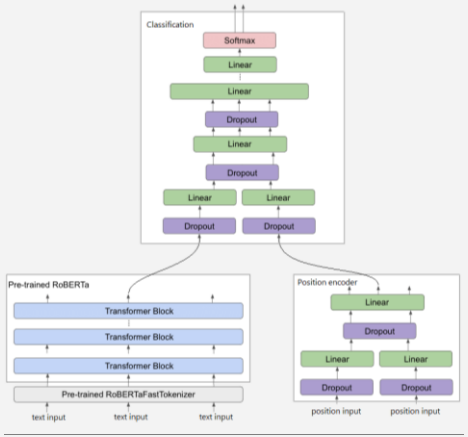
SecureTech Solutions Inc.
123 Cybersecurity Way, Suite 101
TechCity, SecureZone 54321
+34 555 123 456
www.securetechsolutions.com

Aquest correu electrònic i qualsevol fitxer adjunt són confidencials i estan destinats exclusivament a l'ús de la persona o entitat a la qual s'adreça. Si heu rebut aquesta comunicació per error, aviseu immediatament al remitent i elimineu el missatge del vostre sistema. Qualsevol ús, divulgació o distribució no autoritzats està estrictament prohibit.

Si us plau, tingueu precaució quan compartiu aquest correu electrònic amb persones que no necessiten accés al seu contingut. La informació continguda aquí pot ser sensible i subjecta a restriccions legals.

En un esforç per minimitzar l'impacte ambiental i salvaguardar la informació sensible, si us plau, abstenir-vos d'imprimir aquest correu electrònic tret que sigui absolutament necessari.

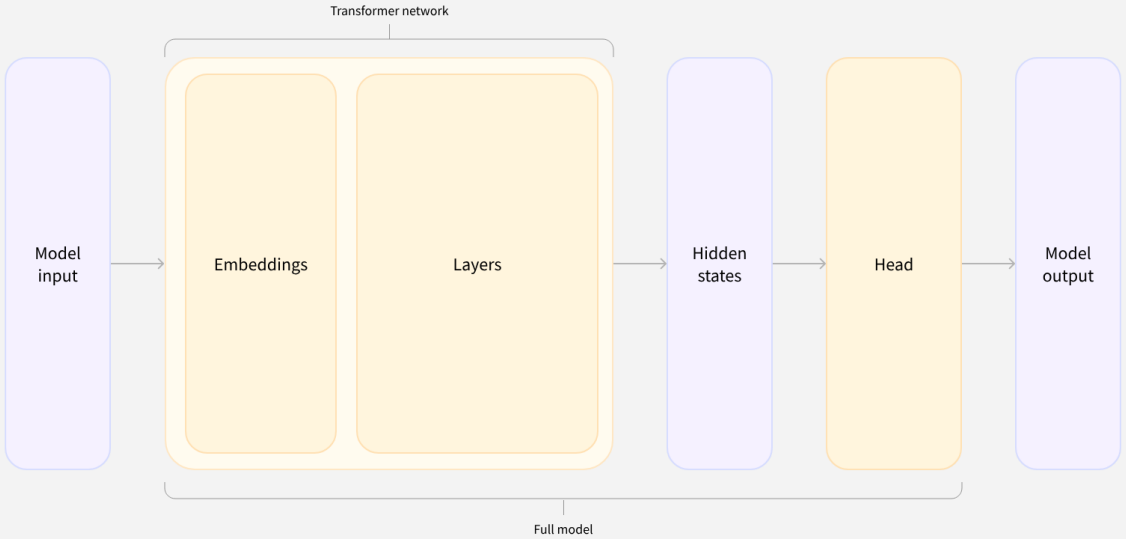
RoBERTa



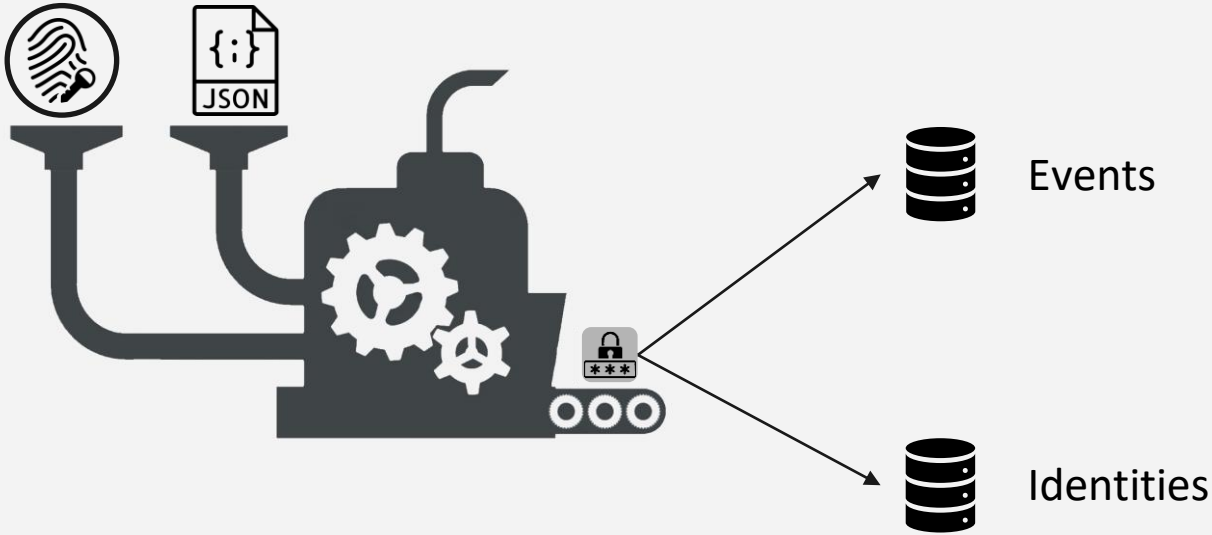
Spacy

Posició	Frase
0	Bon dia,
1	Escric per informar d'un correu...
17	Rebeu una cordial salutació,
18	Departament de ciberseguretat
19	SecureTech Solutions Inc.
22	+34 555 123 456
23	www.securethechsolutions.com
24	Aquest correu electrònic i ...

Execució model NLP



Anonimització



Elasticsearch

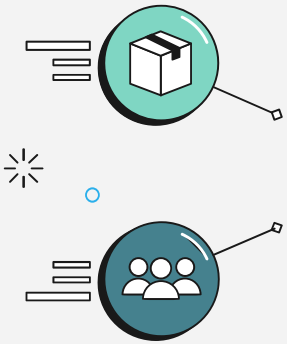
Events

Hash de tots els camps

Identities

Parella clau-valor

API



Processar un tiquet

Processar una llista de tiquets

FastAPI 0.1.0 OAS 3.1

POST /processar_tiquet/ Process Tiquet Endpoint

POST /processar_excel/ Process Excel Endpoint

Parameters

Name	Description
column_number * required integer (query)	column_number

Request body required

file * required
string($binary)

Responses

Code	Description
200	Successful Response
422	Validation Error

04

Entrenament models

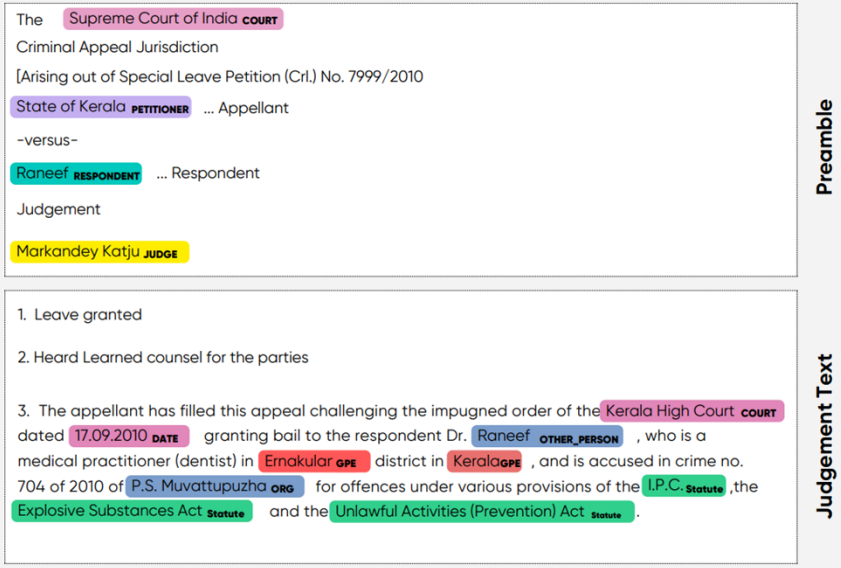
Dades sintètiques

Entitats que depenen del context

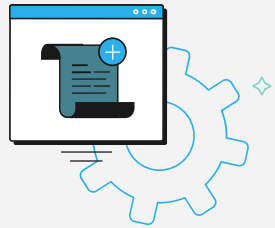
Entitats llargues

Generació de text

Reconeixement d'Entitats a Sentències Judicials de l'Índia



Models destacats



LocalGPT

Mistral 7B

Llama 2

No hi ha suficient memòria

Família Flan-T5

Flan-T5-base
69,36%

Dades reals

- Usuaris afectats
- Accions de mitigació
- Accions de control
- Mail de la víctima
- Mail de l'atacant
- Assumpte del correu
- URL de l'incident

Extracció d'informació en tiquets de phishing

Mètrica	Total
NotFound *test*	72.3%
NotFound *train*	67.6%
Total	68.3%

Models destacats

Família Flan-T5

Mistral 7B

Llama 2

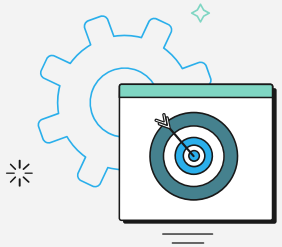
FLOR 6.3B

Qwen 1.5

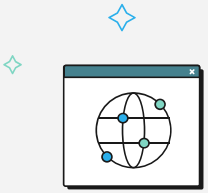
05

Avaluació

Percentatge d’encerts

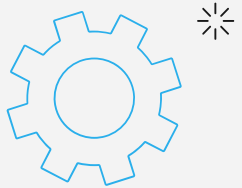


	FLOR 6.3B	QWEN1.5 7B
Usuaris afectats	0%	53,3%
Accions de mitigació	0%	80,0%
Accions de control	0%	93,3%
Mail de la víctima	0%	86,6%
Mail de l’atacant	0%	53,3%
Assumpte del correu	0%	60,0%
URL de l’incident	0%	73,3%
Total	**0%**	**71,4%**



Percentatge de “NotFound”

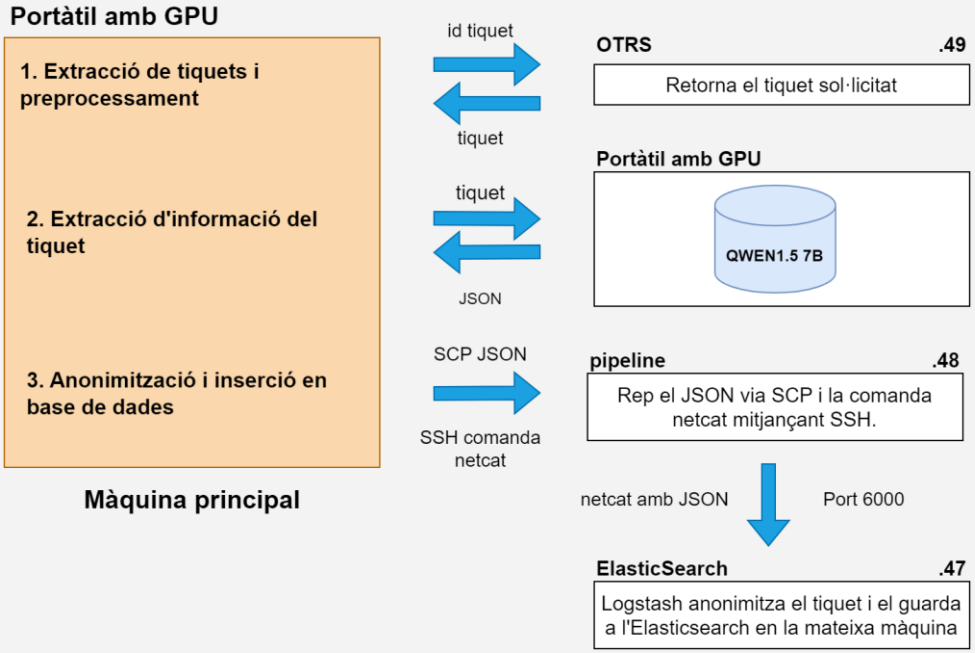
	Conjunt d’entrenament	QWEN1.5 7B
Usuaris afectats	50,0%	80,0%
Accions de mitigació	63,7%	93,3%
Accions de control	97,5%	100,0%
Mail de la víctima	52,5%	80,0%
Mail de l’atacant	56,2%	86,6%
Assumpte del correu	61,2%	86,6%
URL de l’incident	92,5%	86,6%
Total	**67,6%**	**89,5%**



06

Conclusions

Desplegament temporal



Anàlisi automàtica de tiquets de phishing

mitjançant el processament del llenguatge natural

Jaume Casals Vilaplana