
This is the **published version** of the bachelor thesis:

Meloni Martínez, Luca. *On està la meva privacitat en línia? Classificació temàtica de trànsit web xifrat mitjançant aprenentatge automàtic*. Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/334457>

under the terms of the  license.

On està la meva privacitat en línia? Classificació temàtica de trànsit web xifrat mitjançant aprenentatge automàtic

Luca Meloni Martínez

21 de juny de 2026

Resum— Aquest treball estudia la possibilitat d'inferir la categoria temàtica d'una pàgina web a partir de trànsit xifrat, capturat a través de la xarxa Tor, mitjançant tècniques d'aprenentatge automàtic. Per a això, es comparen representacions seqüencials i agregades del trànsit utilitzant models basats en Transformers, XGBoost i una aproximació híbrida. L'avaluació es realitza sobre un conjunt de 66.067 connexions distribuïdes en vuit categories temàtiques. Els resultats mostren que la classificació temàtica és possible a partir de metadades de xarxa, obtenint una precisió Top-1 del 28,50% amb el millor model. A més, s'observa que les representacions agregades superen clarament les aproximacions purament seqüencials i que una part important de la informació discriminativa es concentra en les primeres fases de la connexió. Aquests resultats posen de manifest les implicacions de privacitat associades al Website Fingerprinting.

Paraules clau— Website Fingerprinting, trànsit xifrat, Tor, classificació temàtica, aprenentatge automàtic, XGBoost, Transformers, privacitat.

Abstract— This work studies the feasibility of inferring the thematic category of a website from encrypted traffic, captured through the Tor network, using machine learning techniques. To this end, sequential and aggregated traffic representations are compared using Transformer-based models, XGBoost, and a hybrid approach. The evaluation is performed on a dataset containing 66,067 connections distributed across eight thematic categories. The results show that thematic classification is possible from network metadata, achieving a Top-1 accuracy of 28.50% with the best-performing model. Furthermore, aggregated representations clearly outperform purely sequential approaches, and a significant portion of the discriminative information is concentrated in the early stages of the connection. These findings highlight the privacy implications associated with Website Fingerprinting techniques.

Keywords— Website Fingerprinting, encrypted traffic, Tor, machine learning, traffic classification, network privacy, XGBoost, Transformers.



1 INTRODUCCIÓ

L'ADOPCIÓ generalitzada de protocols de comunicació segurs com HTTPS ha contribuït significativament a millorar la privacitat de les comunicacions a Internet. Tot i això, el xifrat no elimina tota la informació observable associada a una connexió. Features (característiques) com la mida dels paquets, els temps entre transmissions o els patrons de direcció del trànsit continuen sent visibles per a qualsevol observador amb accés al flux de dades i poden contenir informació sobre l'activitat dels usuaris.

Aquest problema resulta especialment rellevant en entorns orientats a la privacitat com la xarxa Tor, utilitzada en aquest treball. Tot i que Tor proporciona mecanismes d'anonimització i xifra les comunicacions entre els diferents nodes de la xarxa, els patrons de trànsit observables continuen sent accessibles i poden contenir informació suficient per inferir propietats sobre les pàgines visitades.

En aquest context apareix el problema conegut com a *Website Fingerprinting*, una família de tècniques que intenta inferir informació sobre les pàgines visitades

a partir dels patrons observables del trànsit de xarxa. Tot i que tradicionalment aquest problema s'ha estudiat per identificar pàgines web concretes, també resulta interessant analitzar si és possible inferir informació més general, com ara la categoria temàtica de les webs visitades.

Aquest treball estudia la viabilitat d'aquesta tasca mitjançant tècniques d'aprenentatge automàtic aplicades a metadades de trànsit de xarxa. Concretament, es comparen diferents representacions del trànsit i diferents famílies de models per determinar fins a quin punt és possible classificar categories temàtiques a partir de trànsit xifrat capturat durant la navegació a través de la xarxa Tor i analitzar les implicacions de privacitat associades.

1.1 Hipòtesi i objectius

La hipòtesi principal d'aquest treball és que és possible inferir informació sobre la categoria temàtica d'una pàgina web a partir del seu trànsit xifrat utilitzant tècniques d'aprenentatge automàtic.

L'objectiu general del treball és avaluar la viabilitat de la classificació temàtica de llocs web a partir de trànsit xifrat i estudiar quines representacions i models resulten més adequats per abordar aquest problema.

Per assolir aquest objectiu general es plantegen els següents objectius específics:

- Analitzar l'estat de l'art en Website Fingerprinting i en tècniques d'anàlisi de trànsit xifrat.
- Estudiar diferents formes de representació del trànsit de xarxa, incloent aproximacions seqüencials i representacions agregades.
- Implementar i avaluar diferents models d'aprenentatge automàtic per a la classificació temàtica de webs.
- Analitzar els factors que influeixen en el rendiment dels models, incloent la selecció de features i la configuració experimental.
- Reflexionar sobre les limitacions i les implicacions de privacitat derivades de la possibilitat d'inferir informació a partir de trànsit xifrat.

2 ESTAT DE L'ART

El Website Fingerprinting és una tècnica d'anàlisi de trànsit que busca inferir informació sobre les pàgines visitades per un usuari a partir de patrons observables en la comunicació, fins i tot quan el contingut es troba xifrat. Treballs pioners van demostrar que característiques com la mida dels paquets, la direcció del flux o els temps entre transmissions poden actuar com una empremta característica del comportament d'una pàgina web [1, 2]. Aquest problema és especialment rellevant en entorns orientats a la privacitat, com la xarxa Tor, on l'objectiu és dificultar la vinculació entre els usuaris i la seva activitat en línia.

Durant els primers anys de recerca, la major part dels sistemes de Website Fingerprinting es basaven en models

d'aprenentatge automàtic tradicionals construïts a partir de features definides manualment. Aquestes aproximacions utilitzaven descriptors estadístics relacionats amb el volum de trànsit, la distribució de mides de paquet o els patrons temporals de les connexions. Tot i obtenir resultats prometedors, diversos treballs van assenyalar que la seva eficàcia podia veure's afectada per canvis en les condicions experimentals i per les limitacions dels conjunts de dades utilitzats [3].

L'aparició del Deep Learning va permetre abordar el problema des d'una perspectiva diferent. En lloc de dependre exclusivament de features definides manualment, aquestes tècniques permeten aprendre representacions directament a partir de les dades observades. Treballs com Deep Fingerprinting [7] o Var-CNN [8] van demostrar que les arquitectures neuronals podien capturar patrons complexos del trànsit i assolir nivells de rendiment superiors als de moltes aproximacions tradicionals. Aquest canvi va consolidar l'ús de tècniques de Deep Learning com una de les principals línies de recerca dins del Website Fingerprinting.

Paral·lelament, els Transformers han esdevingut una de les arquitectures més influents en l'àmbit de l'aprenentatge profund gràcies a la seva capacitat per modelar dependències a llarg termini mitjançant mecanismes d'atenció [6]. Tot i que van ser introduïts originalment en el camp del processament del llenguatge natural, posteriorment s'han aplicat amb èxit a nombrosos problemes relacionats amb dades seqüencials. Aquesta capacitat els converteix en candidats naturals per a l'anàlisi de trànsit de xarxa, ja que una connexió pot entendre's com una seqüència ordenada de paquets.

Diversos treballs recents han explorat l'aplicació de Transformers al Website Fingerprinting. Entre aquests destaca la proposta de Jin et al. [9], que utilitza una arquitectura basada en Transformers per analitzar trànsit web en escenaris amb múltiples pestanyes obertes simultàniament. Aquests resultats suggereixen que els mecanismes d'atenció poden capturar relacions rellevants entre paquets i patrons temporals presents en el trànsit observat.

A diferència de la major part dels treballs existents, centrats principalment en la identificació de dominis web concrets, aquest projecte estudia la classificació de pàgines segons la seva categoria temàtica. Aquest enfocament permet analitzar fins a quin punt el trànsit xifrat conté informació suficient no només per distingir webs individuals, sinó també per inferir propietats més generals sobre l'activitat dels usuaris.

3 METODOLOGIA

3.1 El pipeline

La Figura 1 mostra una visió general del pipeline utilitzat al llarg del treball. El procés es divideix en quatre etapes principals: generació del dataset, representació del trànsit, entrenament dels models i avaluació dels resultats.

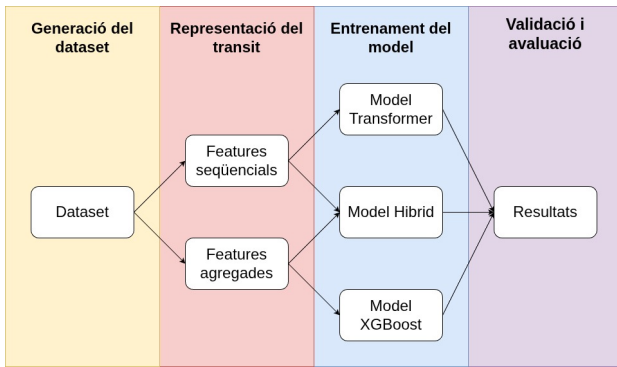


Fig. 1: Esquema del pipeline de treball

A partir de captures de trànsit de xarxa es construeix un dataset estructurat que associa cada connexió amb la seva categoria temàtica. Posteriorment, les connexions es transformen en dues representacions principals: una representació seqüencial basada en la successió temporal de paquets i una representació agregada basada en features estadístiques.

Aquestes representacions s'utilitzen com a entrada dels diferents models estudiats, incloent Transformers, XGBoost i una aproximació híbrida. Finalment, tots els models s'avaluen mitjançant un protocol comú basat en particions independents d'entrenament, validació i test.

3.2 El dataset

Per tal d'entrenar i avaluar els models estudiats en aquest treball, s'ha utilitzat un conjunt de dades format per captures de trànsit web xifrat obtingudes durant la navegació a través de la xarxa Tor i emmagatzemades originalment en format PCAP. Aquestes captures han estat processades i convertides a fitxers CSV mitjançant l'eina Tshark, amb l'objectiu de facilitar-ne l'anàlisi i l'extracció de features. L'ús de Tor permet estudiar el problema en un escenari especialment rellevant dins de la literatura de Website Fingerprinting, orientat a preservar la privacitat dels usuaris.

A diferència dels enfocaments tradicionals de Website Fingerprinting, que acostumen a centrar-se en la identificació de dominis web concrets, aquest treball planteja una tasca de classificació per categories temàtiques. El conjunt de dades està format per vuit categories: bellesa, turisme, bancs, salut, esports, política, ciència i tecnologia.

En total, el dataset conté 66.067 mostres vàlides distribuïdes de manera relativament equilibrada entre les diferents categories. La Figura 2 mostra la distribució de mostres per categoria. Aquesta distribució equilibrada redueix el risc que els models desenvolupats afavoreixin determinades classes únicament per la seva major presència al conjunt de dades i facilita una comparació més representativa del rendiment entre categories.

A més de la classificació temàtica, el dataset conserva informació sobre el domini web associat a cada mostra. En conjunt, les dades inclouen 163 dominis únics agrupats dins les diferents categories, fet que permet estudiar fins a quin

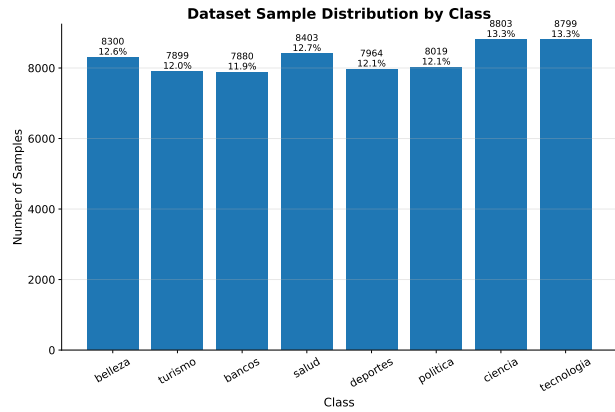


Fig. 2: Distribució de mostres per categoria temàtica.

punt existeixen patrons de trànsit compartits entre webs d'una mateixa temàtica. Aquesta característica resulta especialment rellevant, ja que l'objectiu del treball no és identificar pàgines concretes sinó determinar si existeixen comportaments comuns associats a categories més generals.

L'estructura del dataset segueix una organització jeràrquica basada en tres nivells principals: categoria, lloc web i mostra. La Figura 3 mostra de manera esquemàtica aquesta organització. Cada mostra correspon a un fitxer CSV individual derivat d'una captura de trànsit i constitueix la unitat bàsica emprada durant l'entrenament, la validació i l'avaluació dels models.

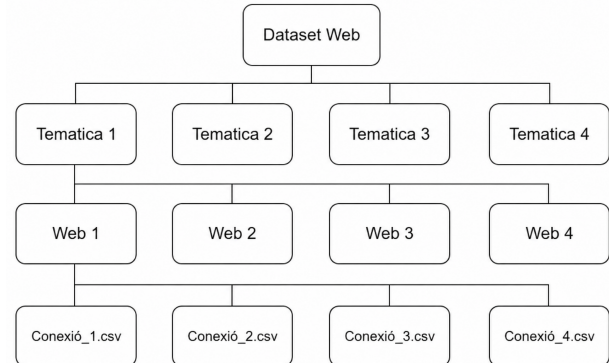


Fig. 3: Estructura jeràrquica del dataset.

Aquesta estructura permet mantenir la traçabilitat entre categories, dominis i captures individuals, facilitant tant la preparació dels experiments com l'anàlisi posterior dels resultats.

3.3 Representació del trànsit

Els fitxers CSV generats a partir de les captures PCAP contenen únicament metadades del trànsit observat, incloent l'instant temporal de captura, l'adreça IP d'origen, la longitud del paquet i el protocol utilitzat.

A partir d'aquesta informació es van construir dues representacions diferents de cada connexió. La primera correspon a una representació seqüencial, en què cada connexió es modela com una successió ordenada de paquets descrits mitjançant features temporals, direccionals i de volum de trànsit. La segona consisteix en una representació

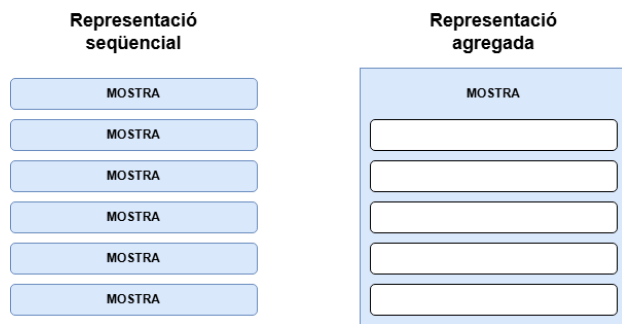


Fig. 4: Esquema del model seqüencial vs agregat

agregada basada en estadístiques globals de la connexió, com la durada, el nombre de paquets, el volum de trànsit o la distribució dels temps entre paquets.

La utilització de dues representacions diferents respon als requisits dels models estudiats. Mentre que els Transformers aprofiten la informació temporal present en les seqüències de paquets, models tabulars com XGBoost treballen de manera més eficient amb vectors de features de longitud fixa. Aquesta separació permet comparar diferents famílies de models utilitzant representacions adequades per a cadascuna d'elles.

3.4 Els models

Amb l'objectiu d'avaluar diferents estratègies per a la classificació temàtica de trànsit web xifrat, es van estudiar tres aproximacions basades en paradigmes d'aprenentatge diferents: un model Transformer, un model XGBoost i una aproximació híbrida que combina informació seqüencial i agregada. Aquesta selecció permet comparar models especialitzats en el processament de seqüències amb models orientats a dades tabulars, així com explorar si la combinació d'ambdues representacions pot aportar informació complementària.

3.4.1 Transformer

El Transformer va constituir el punt de partida del projecte i representa l'aproximació més alineada amb la hipòtesi inicial del treball. En aquest cas, el model rep com a entrada la representació seqüencial de cada connexió, modelada com una successió ordenada de paquets.

La motivació principal per utilitzar aquesta arquitectura és la seva capacitat per capturar dependències temporals i patrons complexos presents en el trànsit de xarxa mitjançant mecanismes d'atenció. Aquesta propietat la converteix en una alternativa especialment interessant per a problemes de Website Fingerprinting, on l'ordre dels paquets pot contenir informació rellevant per a la classificació.

3.4.2 XGBoost

XGBoost es va incorporar com a alternativa basada en dades tabulars. Aquest algorisme es basa en conjunts d'arbres de decisió entrenats de manera iterativa i és especialment adequat per a problemes de classificació sobre vectors de

features de longitud fixa.

A diferència del Transformer, XGBoost treballa sobre la representació agregada de les connexions, on cada mostra es resumeix mitjançant un conjunt de features estadístiques que descriuen el comportament global del trànsit. Aquesta representació permet concentrar informació rellevant de tota la connexió en un únic vector, facilitant el procés d'aprenentatge.

3.4.3 Aproximació híbrida

A més dels dos enfocaments principals, es va estudiar una aproximació híbrida que combina informació seqüencial i agregada. El model incorpora una branca encarregada de processar la seqüència de paquets i una segona branca que rep features agregades de la connexió, fusionant ambdues representacions abans de la classificació final.

L'objectiu d'aquesta aproximació és analitzar si les dues representacions aporten informació complementària i si la seva combinació permet millorar el rendiment respecte als models que utilitzen únicament una de les dues fonts d'informació.

3.5 Estratègia d'avaluació i validació

Per tal de garantir una avaluació rigorosa, els experiments es van dur a terme utilitzant particions independents d'entrenament, validació i test, generades de manera estratificada segons la categoria temàtica.

El conjunt d'entrenament es va utilitzar per ajustar els models, mentre que el conjunt de validació es va emprar durant la selecció d'hiperparàmetres i configuracions experimentals. Finalment, el conjunt de test es va reservar exclusivament per a l'avaluació final dels models seleccionats.

La comparació entre models es va basar principalment en les mètriques Top-1 Accuracy i Macro-F1. Mentre que la Top-1 Accuracy mesura el percentatge de mostres classificades correctament, la Macro-F1 proporciona una avaluació equilibrada entre categories.

L'ús de particions independents va resultar fonamental per detectar possibles fonts de biaix experimental i validar les diferents configuracions estudiades, permetent adoptar una metodologia consistent per a totes les avaluacions realitzades.

4 RESULTATS I DISCUSSIÓ

4.1 Resultats generals

La Taula 1 resumeix els resultats finals de les tres aproximacions principals estudiades: el Transformer, el model híbrid i XGBoost. La comparació es basa principalment en la Top-1 Accuracy sobre el conjunt de test, ja que aquesta mètrica indica directament el percentatge de mostres classificades correctament. Addicionalment, també es reporta la Macro-F1, que proporciona una visió més equilibrada

Model	Representació	Top-1	Macro-F1
Transformer	Seqüencial	16.60%	14.85%
Hybrid CNN	Híbrida	26.43%	25.58%
XGBoost	Agregada	28.50%	27.93%

TAULA 1: TAULA COMPARATIVA DELS RESULTATS

del rendiment entre totes les categories.

Els resultats mostren una diferència clara entre les aproximacions estudiades. El Transformer, basat exclusivament en la representació seqüencial dels paquets, obté un Top-1 Accuracy del 16,60% i una Macro-F1 del 14,85%. La incorporació d'informació agregada en el model híbrid produeix una millora substancial, assolint un Top-1 Accuracy del 26,43% i una Macro-F1 del 25,58%.

El millor rendiment global s'obté amb XGBoost sobre la representació agregada seleccionada, que aconsegueix un Top-1 Accuracy del 28,50% i una Macro-F1 del 27,93%. Tot i que la diferència respecte al model híbrid és moderada, els resultats mostren de manera consistent que les aproximacions basades en features agregades superen clarament el model seqüencial pur.

Resulta especialment rellevant observar que la millora més important no es produeix entre el model híbrid i XGBoost, sinó entre el Transformer i el model híbrid. Aquest comportament suggereix que la incorporació de descriptors agregats aporta molta més informació discriminativa que la substitució d'una arquitectura per una altra. En conseqüència, els resultats apunten que la representació de les dades té un paper determinant en aquest problema, fins i tot per sobre de l'elecció del model utilitzat.

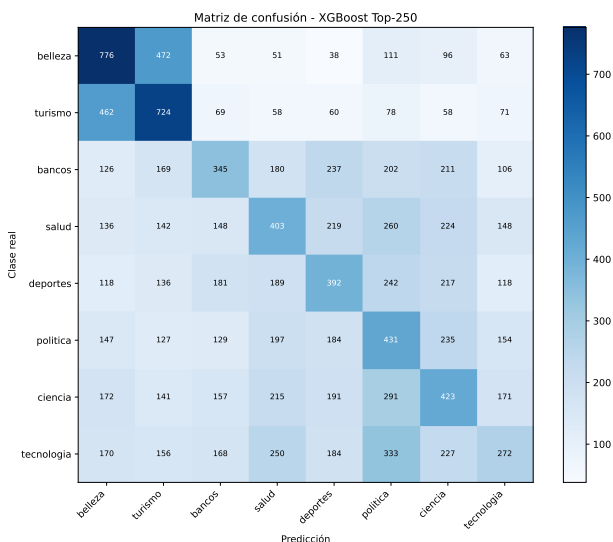


Fig. 5: Matriu de confusió corresponent al millor model obtingut en aquest treball.

La Figura 5 mostra la matriu de confusió corresponent al millor model obtingut. Es pot observar que dues categories concentren una part important dels encerts, mentre que la resta presenten un patró de confusió més distribuït. Aquest comportament indica que algunes categories generen

patrons de trànsit més distintius que altres i suggereix que la dificultat del problema no és homogènia per a totes les classes.

En conjunt, aquests resultats suggereixen que la informació discriminativa necessària per diferenciar categories temàtiques es troba més ben representada mitjançant descriptors agregats de la connexió que no pas exclusivament a través de la seqüència de paquets observada.

4.2 Anàlisi de les features

Per tal d'identificar quins aspectes del trànsit aportaven més informació per a la classificació, es va analitzar la importància de les features utilitzades pel model XGBoost. Partint d'una representació agregada formada inicialment per 1059 variables, les features es van ordenar segons les mesures d'importància proporcionades pel model durant l'entrenament. Aquest procediment va permetre generar un rànquing de features i construir subconjunts reduïts utilitzats durant el procés de selecció de variables.

TAULA 2: TOP-10 FEATURES MÉS IMPORTANTS SEGONS EL MODEL XGBOOST.

Pos.	Feature	Família
1	first_10_delta_t_p10	Timing inicial
2	out_delta_t_min	Timing dir.
3	out_packet_len_std	Mida de paquet
4	in_packet_len_p90	Mida de paquet
5	out_packet_len_p90	Mida de paquet
6	first_5_packet_len_p25	Paquets inicials
7	first_10_packet_len_p10	Paquets inicials
8	out_delta_t_p10	Timing dir.
9	log_delta_t_p50	Timing
10	out_in_delta_t_median_ratio	Direcció

La Taula 2 mostra les deu features més importants del model final. L'anàlisi del rànquing revela que la informació discriminativa es concentra principalment en quatre famílies de descriptors: features temporals, estadístiques de mida de paquet, informació direccional i mesures relacionades amb els bursts de trànsit.

Les features temporals presenten una presència especialment destacada. La variable amb major importància és `first_10_delta_t_p10`, corresponent al percentil 10 dels temps entre paquets observats durant els primers deu paquets de la connexió. Altres features com `out_delta_t_min`, `out_delta_t_p10`, `first_150_delta_t_p10` o `first_300_delta_t_p10` també apareixen entre les primeres posicions del rànquing.

Les estadístiques de mida de paquet constitueixen una segona família especialment rellevant. Features com `out_packet_len_std`, `in_packet_len_p90` o `out_packet_len_p90` indiquen que la distribució de les mides dels paquets aporta informació útil per diferenciar categories. De manera similar, la presència de features relacionades amb els bursts inicials suggereix que l'organització del trànsit durant les primeres fases de la connexió

conté patrons discriminatius.

Un dels resultats més interessants és que moltes de les features millor classificades es calculen sobre els primers paquets de la connexió. Aquest comportament suggereix que una part significativa de la informació necessària per diferenciar categories es troba disponible durant les fases inicials de la comunicació. També destaca la presència simultània de features temporals, de mida i direccionals entre les primeres posicions del rànquing, indicant que el model combina informació de naturalesa diversa en lloc de dependre d'una única propietat del trànsit.

Finalment, el fet que subconjunts reduïts de features mantinguin un rendiment molt proper al model complet suggereix que gran part de la informació útil es concentra en un nombre relativament petit de descriptors, evidenciant una certa redundància entre les variables generades inicialment.

4.3 Anàlisi dels hiperparàmetres

A més de l'enginyeria de features, també es va estudiar l'impacte de l'ajust d'hiperparàmetres sobre el rendiment del model XGBoost. Per fer-ho, es van explorar diferents configuracions variant la profunditat dels arbres (`max_depth`), el nombre d'estimadors (`n_estimators`), la taxa d'aprenentatge (`learning_rate`) i els paràmetres de mostreig (`subsample` i `colsample_bytree`).

Els resultats mostren que l'ajust d'hiperparàmetres permet obtenir millores modestes sobre la configuració inicial. La millor configuració va aconseguir una Top-1 Accuracy de validació del 27,86%, representant una millora de 0,47 punts percentuals respecte al model de referència. Les configuracions millor classificades comparteixen profunditats moderades, un nombre elevat d'estimadors i taxes d'aprenentatge reduïdes.

Tanmateix, aquesta millora no es va traduir en un augment equivalent del rendiment sobre test. Després de reentrenar la millor configuració utilitzant conjuntament les dades d'entrenament i validació, el model va obtenir una Top-1 Accuracy del 28,15%, lleugerament inferior al 28,50% del model de referència.

Aquests resultats suggereixen que el rendiment final està més condicionat per la representació de les dades i pel procés de selecció de features que no pas per l'ajust fi dels hiperparàmetres. En aquest context, l'ajust d'hiperparàmetres ha actuat principalment com un mecanisme de refinament, mentre que les millores més significatives provenen de les decisions relacionades amb la representació del trànsit.

4.4 Discussió dels resultats

Els resultats obtinguts permeten extreure diverses conclusions rellevants sobre la classificació temàtica de trànsit web xifrat i sobre les diferents aproximacions estudiades al llarg del treball.

La classificació temàtica és possible a partir de trànsit xifrat.

El primer resultat destacable és que tots els models estudiats obtenen un rendiment superior a l'esperable per atzar. Aquest fet suggereix que existeix una correlació entre les característiques observables del trànsit i la categoria temàtica de la pàgina visitada. Per tant, fins i tot sense accedir al contingut de les comunicacions, és possible extreure informació parcial sobre l'activitat de l'usuari a partir de metadades de xarxa.

Aquesta conclusió també es reflecteix en la matriu de confusió, on s'observa que determinades categories són identificades amb una precisió considerablement superior a la resta. Per contra, la major part dels errors es produeixen entre categories amb patrons de trànsit similars, fet que suggereix que la informació discriminativa no es distribueix de manera uniforme entre totes les classes.

La representació és més important que el model.

La comparació entre les diferents aproximacions mostra que el factor més determinant no és tant l'arquitectura utilitzada com la representació de les dades. Tant el model híbrid com XGBoost, que incorporen informació agregada de la connexió, superen clarament el Transformer basat exclusivament en seqüències de paquets. Aquest resultat suggereix que, en el context del dataset estudiat, les estadístiques agregades capturen millor la informació discriminativa necessària per diferenciar categories.

Aquest resultat és especialment rellevant perquè el Transformer era, en principi, l'arquitectura amb una major capacitat per modelar dependències complexes dins de les seqüències de paquets. Tot i això, els resultats mostren que aquesta capacitat no es tradueix necessàriament en un millor rendiment quan la representació de les dades no captura de manera eficient la informació més discriminativa del problema.

La millora observada en passar del Transformer al model híbrid és considerablement superior a la diferència existent entre el model híbrid i XGBoost. Aquest comportament suggereix que una part important de la informació necessària per a la classificació ja es troba recollida en les features agregades de la connexió. En conseqüència, els resultats indiquen que les decisions relacionades amb la representació de les dades poden tenir un impacte superior a les relacionades exclusivament amb l'elecció de l'arquitectura.

La informació més rellevant es concentra en les fases inicials de la connexió.

L'anàlisi de les features més importants mostra una presència recurrent de variables relacionades amb els primers paquets, els primers bursts i els primers intervals temporals de la connexió. Aquest comportament suggereix que una part significativa de la informació utilitzada pels models es troba disponible molt aviat, abans que la connexió hagi transferit una quantitat significativa de dades.

Aquest resultat és coherent amb la naturalesa de moltes connexions web modernes. Les primeres fases de la comunicació acostumen a concentrar processos com l'establiment de la connexió i les primeres respostes del servidor, elements que poden variar significativament entre diferents tipus de webs. El fet que moltes de les features més importants estiguin associades a aquestes primeres interaccions suggereix que una part rellevant de la informació discriminativa es troba disponible molt abans que la connexió hagi finalitzat.

Els resultats depenen del context experimental.

Tot i que els resultats obtinguts mostren l'existència d'informació discriminativa en el trànsit observat, aquests s'han d'interpretar dins del context experimental utilitzat. Els experiments es van realitzar sobre particions estratificades per categoria i no constitueixen una demostració directa de generalització a webs completament noves o a captures obtingudes en entorns diferents.

Aquesta limitació és especialment important en problemes relacionats amb Website Fingerprinting, on petites diferències en el procés de captura o en la composició del conjunt de dades poden influir de manera significativa en el rendiment observat. Per aquest motiu, els resultats presentats s'han d'entendre com una evidència de la viabilitat de la classificació temàtica en el context estudiat, més que no pas com una demostració definitiva de la seva capacitat de generalització a qualsevol escenari.

5 CONCLUSIONS

5.1 Conclusions principals

Aquest treball ha estudiat la possibilitat d'inferir la categoria temàtica d'una pàgina web a partir de metadades de trànsit xifrat. Els resultats obtinguts mostren que aquesta tasca és viable, ja que tots els models avaluats aconseguixen resultats significativament superiors als esperables per atzar. Aquest fet confirma l'existència d'informació discriminativa en el trànsit observat i posa de manifest que el xifrat del contingut no elimina completament la possibilitat d'inferir informació sobre l'activitat de l'usuari.

Una de les principals conclusions del treball és que la representació de les dades té un impacte superior al de l'arquitectura utilitzada. Els millors resultats s'han obtingut amb XGBoost sobre features agregades, superant clarament les aproximacions basades exclusivament en seqüències de paquets. L'anàlisi de les features més rellevants indica, a més, que una part important de la informació discriminativa es concentra en les primeres fases de la connexió.

Finalment, tot i que els resultats obtinguts són consistents amb la hipòtesi plantejada, la generalització d'aquestes conclusions a nous escenaris continua sent una qüestió oberta. Per aquest motiu, els resultats s'han d'interpretar dins del context experimental utilitzat i constitueixen un punt de partida per a futures línies d'investigació en l'àmbit del Website Fingerprinting i de l'anàlisi de trànsit xifrat.

5.2 Implicacions

Els resultats obtinguts en aquest treball tenen implicacions rellevants des del punt de vista de la privacitat. Tot i que el trànsit analitzat es correspon amb connexions realitzades a través de la xarxa Tor i es troba xifrat, els experiments mostren que les seves característiques observables continuen contenint senyals útils per inferir propietats de l'activitat de l'usuari.

En particular, la capacitat dels models per distingir categories temàtiques amb una precisió significativament superior a l'atzar suggereix que un observador amb accés al trànsit podria obtenir coneixement parcial sobre els hàbits de navegació dels usuaris sense necessitat de desxifrar les comunicacions. Aquest tipus d'inferències podria utilitzar-se per identificar interessos generals o construir perfils d'ús basats exclusivament en metadades de xarxa.

Tot i que els resultats presentats no constitueixen una demostració d'un atac pràctic en entorns reals, sí que evidencien que el xifrat del contingut, per si sol, no garanteix l'ocultació completa de tota la informació associada a una connexió. Per aquest motiu, la protecció de la privacitat en sistemes de comunicació moderns continua sent un repte obert que requereix mecanismes addicionals capaços de reduir la informació exposada pels patrons de trànsit observables.

5.3 Limitacions actuals

Tot i que els resultats obtinguts són consistents amb la hipòtesi plantejada, aquest treball presenta diverses limitacions que cal tenir en compte a l'hora d'interpretar-ne les conclusions.

En primer lloc, els experiments s'han realitzat sobre un únic dataset construït a partir d'un conjunt concret de pàgines web agrupades en categories temàtiques. Això implica que els resultats reflecteixen el comportament dels models dins del context específic estudiat i no permeten garantir automàticament la seva aplicabilitat a altres col·leccions de webs o entorns de navegació. Per aquest motiu, una línia de treball futur rellevant consisteix a avaluar els models sobre nous conjunts de dades i condicions de captura que permetin estudiar amb més profunditat la seva capacitat de generalització.

En segon lloc, durant el desenvolupament del projecte es van identificar possibles fonts de biaix associades al procés de captura i preprocessament de les dades. Tot i que es van implementar mesures per reduir aquests efectes i es va validar el protocol experimental utilitzat, no es pot descartar completament que determinades característiques del dataset contribueixin parcialment al rendiment observat. En aquest sentit, futurs treballs podrien explorar protocols d'avaluació més exigents orientats a reduir la influència de possibles artefactes experimentals.

Una limitació addicional és que les particions utilitzades es van generar de manera estratificada per categoria, però no es va imposar una separació estricta per dominis

web. Com a conseqüència, és possible que els models aprenguin parcialment patrons associats a determinades webs presents tant en entrenament com en avaluació. Una línia de treball especialment interessant consistiria a dissenyar experiments basats en dominis completament no vistos durant l'entrenament, permetent així estudiar de manera més rigorosa la capacitat dels models per generalitzar a noves pàgines dins d'una mateixa categoria temàtica.

Una altra limitació és que només s'han estudiat dues representacions principals del trànsit: la representació seqüencial i la representació agregada. Els resultats suggereixen que les features agregades contenen una gran quantitat d'informació discriminativa, però també indiquen que existeix marge per desenvolupar noves representacions capaces de combinar informació seqüencial i estadística de manera més eficient. De manera similar, també seria interessant ampliar l'estudi a altres arquitectures d'aprenentatge automàtic.

Finalment, aquest treball aporta evidència sobre la possibilitat d'inferir informació temàtica a partir de trànsit xifrat, però no constitueix una demostració completa d'un atac realista en entorns de producció. Per aquest motiu, resulta especialment interessant continuar investigant tant les implicacions de privacitat associades a aquests resultats com els mecanismes de defensa capaços de reduir la informació exposada pels patrons de trànsit observables. En particular, futures línies de recerca podrien estudiar l'efectivitat de tècniques d'ofuscació del trànsit, generació de trànsit de cobertura o modificació dels patrons temporals de les connexions amb l'objectiu de dificultar la inferència d'informació a partir de metadades de xarxa.

ÚS DE LA IA GENERATIVA

Durant el desenvolupament del projecte s'han utilitzat eines d'IA generativa com a suport complementari en diferents tasques relacionades amb l'aprenentatge, la programació i la redacció de la memòria.

Principalment, aquestes eines s'han utilitzat per resoldre dubtes puntuals relacionats amb conceptes tècnics de Machine Learning i models Transformer, cercar recursos o documentació rellevant, i ajudar en aspectes de redacció, estil i claredat expositiva de la memòria.

També s'han utilitzat de manera puntual eines d'assistència a la programació per agilitzar tasques tècniques i de depuració de codi durant el desenvolupament dels experiments.

El disseny dels experiments, la definició de la metodologia, l'anàlisi dels resultats i les conclusions del treball han estat desenvolupats i validats manualment durant el projecte.

AGRAÏMENTS

M'agradaria expressar el meu sincer agraïment al meu tutor per la seva disponibilitat, orientació i suport al llarg de tot el desenvolupament d'aquest treball. La seva ajuda ha estat fonamental per resoldre dubtes i superar les dificultats que han anat sorgint durant el projecte. Li agraeixo especialment l'oportunitat d'apropar-me al món de la recerca i de descobrir una mostra de la manera de treballar en aquest àmbit, una experiència que ha reforçat el meu interès per orientar la meua trajectòria professional en aquesta direcció.

També vull agrair a la meua família el suport incondicional, la confiança i la paciència que m'han brindat durant tota la meua etapa universitària. Sense el seu recolzament constant, aquest camí hauria estat molt més difícil.

Finalment, agraeixo als meus amics i companys el suport rebut durant aquests anys. En especial, vull reconèixer aquells que han contribuït al desenvolupament d'aquest projecte aportant idees, suggeriments i ajuda sempre que ha estat necessari. Els seus comentaris i el seu suport han estat de gran valor al llarg de tot el procés.

REFERÈNCIES

- [1] Panchenko, A., Niessen, L., Zinnen, A., & Engel, T. *Website Fingerprinting in Onion Routing Based Anonymization Networks*. Proceedings of the ACM Workshop on Privacy in the Electronic Society (WPES), 2011.
- [2] Wang, T., Cai, X., Nithyanand, R., Johnson, R., & Goldberg, I. *Effective Attacks and Provable Defenses for Website Fingerprinting*. USENIX Security Symposium, 2014.
- [3] Juarez, M., Afroz, S., Acar, G., Diaz, C., & Greensadt, R. *A Critical Evaluation of Website Fingerprinting Attacks*. Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS), 2014.
- [4] LeCun, Y., Bengio, Y., & Hinton, G. *Deep Learning*. Nature, vol. 521, no. 7553, pp. 436–444, 2015.
- [5] Chen, T., & Guestrin, C. *XGBoost: A Scalable Tree Boosting System*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016.
- [6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L.L. & Polosukhin, I. *Attention Is All You Need*. Advances in Neural Information Processing Systems (NeurIPS), 2017.
- [7] Sirinam, P., Imani, M., Juarez, M., & Wright, M. *Deep Fingerprinting: Undermining Website Fingerprinting Defenses with Deep Learning*. Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS), 2018.
- [8] Bhat, S., Lu, D., Kwon, A., & Devadas, S. *Var-CNN: A Data-Efficient Website Fingerprinting Attack Based*

on Deep Learning. Proceedings on Privacy Enhancing Technologies (PoPETs), vol. 2019, no. 1, pp. 292–310, 2019.

- [9] Jin, Z., Lu, T., Luo, S., & Shang, J. *Transformer-Based Model for Multi-Tab Website Fingerprinting Attack*. Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS), 2023.

APÈNDIX

A.1 Anàlisi de les característiques més rellevants

Per tal d'interpretar millor el comportament del model XGBoost seleccionat, es va realitzar una anàlisi de les característiques amb major importància segons el model entrenat amb les 250 variables més rellevants.

Les característiques es van agrupar en diferents famílies segons la informació que representen. El criteri utilitzat va ser el següent:

- Característiques relacionades amb els primers intervals temporals de la connexió (*first*_delta_t*), duracions inicials i taxes inicials: **Timing inicial**.
- Característiques *delta_t* globals o específiques del trànsit entrant o sortint: **Timing**.
- Característiques relacionades amb els primers paquets de la connexió: **Paquets inicials**.
- Característiques basades en longituds de paquets: **Mida de paquet**.
- Característiques derivades de ràfegues de trànsit: **Bursts**.
- Característiques basades en trànsit entrant i sortint: **Direccionalitat**.
- Característiques globals de volum: **Volum global**.

La Taula 3 resumeix la distribució de les 50 característiques més importants segons aquestes famílies.

TAULA 3: DISTRIBUCIÓ PER FAMÍLIES DE LES 50 CARACTERÍSTIQUES MÉS IMPORTANTS.

Família	Nombre de característiques	Percentatge
Timing inicial	15	30%
Timing	10	20%
Paquets inicials	9	18%
Mida de paquet	7	14%
Bursts	6	12%
Direccionalitat	2	4%
Volum global	1	2%
Altres	0	0%
Total	50	100%

La Taula 4 mostra les 20 característiques més importants juntament amb la família a la qual pertany cadascuna.

TAULA 4: FAMÍLIES DE LES 20 CARACTERÍSTIQUES MÉS IMPORTANTS DEL MODEL TOP-250.

Posició	Característica	Família
1	<i>first_10_delta_t_p10</i>	Timing inicial
2	<i>out_delta_t_min</i>	Timing
3	<i>out_packet_len_std</i>	Mida de paquet
4	<i>in_packet_len_p90</i>	Mida de paquet
5	<i>out_packet_len_p90</i>	Mida de paquet
6	<i>first_5_packet_len_p25</i>	Paquets inicials
7	<i>first_10_packet_len_p10</i>	Paquets inicials
8	<i>out_delta_t_p10</i>	Timing
9	<i>log_delta_t_p50</i>	Timing
10	<i>out_in_delta_t_median_ratio</i>	Timing
11	<i>first_5_out_burst_len_std</i>	Bursts
12	<i>first_200_delta_t_p10</i>	Timing inicial
13	<i>in_packet_len_mean</i>	Mida de paquet
14	<i>total_packets</i>	Volum global
15	<i>first_10_out_burst_len_p90</i>	Bursts
16	<i>delta_t_p50</i>	Timing
17	<i>first_150_delta_t_p10</i>	Timing inicial
18	<i>first_20_packet_len_p10</i>	Paquets inicials
19	<i>first_300_delta_t_p10</i>	Timing inicial
20	<i>in_packet_len_max</i>	Mida de paquet

Interpretació

Les famílies relacionades amb el temps dominen clarament el conjunt de característiques més rellevants. Les variables de *Timing inicial* i *Timing* representen conjuntament el 50% de les 50 característiques amb major importància, fet que suggereix que els patrons temporals del trànsit constitueixen la principal font d'informació per a la classificació temàtica.

Adicionalment, les característiques associades als primers paquets de la connexió representen un 18% del total, reforçant la hipòtesi que les fases inicials de càrrega d'una pàgina web contenen una part significativa de la informació discriminativa. Les variables relacionades amb la mida dels paquets i amb els bursts també contribueixen de manera rellevant, mentre que les característiques basades exclusivament en volum global o direccionalitat tenen una presència molt més reduïda.

Aquests resultats són coherents amb les conclusions obtingudes a la discussió, on es va observar que la informació més útil per a la classificació es concentra principalment en els primers instants de la connexió i en els patrons temporals del trànsit.

A.2 Configuracions experimentals

Per tal de facilitar la reproducció dels experiments realitzats, la Taula 5 resumeix les principals configuracions utilitzades pels tres models avaluats:

TAULA 5: CONFIGURACIONS FINALS DELS MODELS PRINCIPALS.

Aspecte	Transformer	XGBoost Top-250	Híbrid
Representació	Seqüencial	Agregada	Seqüencial + agregada
Entrada	1000 paquets, 7 features/paquet	250 features	1000 paquets + 75 features
Split	80/10/10	70/10/20	70/10/20
Criteri de selecció	Val. Top-1	Val. Top-1	Val. Top-1
Learning rate	10^{-4}	0.03	10^{-3}
Batch size	8	–	32
Entrenament	200 èpoques màx.	700 estimadors	100 èpoques màx.
Arquitectura principal	1 capa, 2 heads, dim. 32	Profunditat 5	CNN residual + MLP
Regularització	Dropout 0.10, weight decay 10^{-4}	Subsample 0.75, colsample 0.9	Dropout 0.20, weight decay 10^{-4}
Altres	Seed 42	Seed 42	Seed 42

A.3 Resultats addicionals de selecció de característiques

La Taula 6 mostra els resultats obtinguts durant el procés de selecció de característiques per al model XGBoost. Es van avaluar diferents valors de top_k utilitzant les característiques amb major importància segons el model entrenat.

TAULA 6: RESULTADOS DEL EXPERIMENTO DE SELECCIÓN DE FEATURES XGBOOST

top_k	Val. Top-1	Val. Top-5	Test Top-1	Test Top-5	Configuració seleccionada
25	22.90%	78.99%	–	–	regularized_d5_e700_lr003
50	25.96%	81.23%	–	–	shallower_d4_e900_lr003
75	26.50%	82.22%	–	–	shallower_d4_e900_lr003
100	27.05%	82.62%	–	–	cols_d5_e700_lr003
150	26.65%	82.81%	–	–	regularized_d5_e700_lr003
250	27.40%	83.47%	28.50%	83.67%	regularized_d5_e700_lr003
500	27.36%	83.17%	–	–	regularized_d5_e700_lr003
1059	27.06%	82.94%	–	–	regularized_d5_e700_lr003

Els resultats mostren que l'ús de les 250 característiques més rellevants va proporcionar el millor rendiment en validació i va ser la configuració final seleccionada per a l'avaluació sobre el conjunt de test.