

Classificazione e Predizione del traffico di rete generato dalle applicazioni mobili per l'interazione sociale e lo smart-working



GIORNATA DI INCONTRO
BORSE DI STUDIO GARR
"ORIO CARLINI"
ROMA



DIETI • **UNIVERSITA'** DEGLI STUDI DI
NAPOLI FEDERICO II

Idio Guarino, PhD Student @UniNa

Tutor: Prof. Antonio Pescapè, Professore Ordinario @ DIETI-UNINA



Gli effetti del lockdown

*“il lockdown ha causato una crescita improvvisa dei volumi di traffico, perlopiù legata all’impiego di applicazioni per l’**interazione sociale e lo Smart-working (ISSW)**”*



Gli effetti del lockdown

Feb → Apr 2020, Mondo [1]

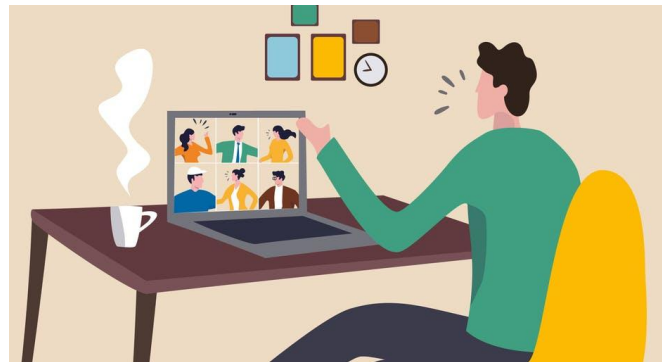
- **+40%** - traffico residenziale (**+121%** Upstream, **+23%** Downstream)

Mar → Giu 2020, Europa [2]

- **+200%** - traffico residenziale applicazioni **ISSW**
- **-55%** - traffico delle reti “educational”

Mar → Apr 2020, Italia - PoliTo **e-teaching** [3]

- fino a **1.5 Gbit/s**
- **> 600** classi virtuali giornaliere
- **> 16k** studenti



1. Sandvine, The Global Internet Phenomena Report COVID-19 Spotlight, 2020

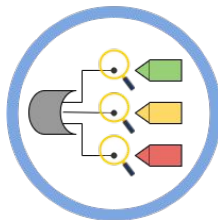
2. A. Feldmann et al., *The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic*, IMC 2020

3. T. Favale et al., *Campus traffic and e-Learning during COVID-19 pandemic*, Elsevier Computer Networks, 2020

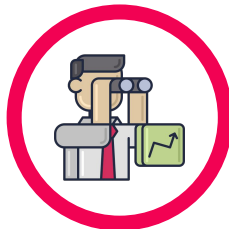


Classificazione e Predizione del traffico

Traffic Classification (TC): Cosa sta attraversando la mia rete?



Traffic Prediction (TP): Come evolveranno le caratteristiche del traffico?





Classificazione e Predizione del traffico

Scenari Applicativi



Quality of Service



Ingegnerizzazione
del traffico



Gestione delle Risorse



Sicurezza

Problemi



Molte App



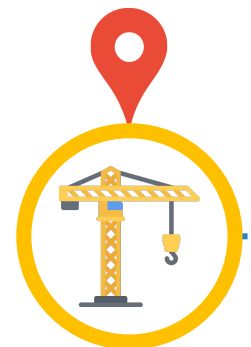
Ambiguità dei servizi di terze parti
Eterogeneità del Traffico (*utente, dispositivo, OS*)
Dinamicità ed evoluzione continua



Crittografia
NAT
Dynamic Port Mapping



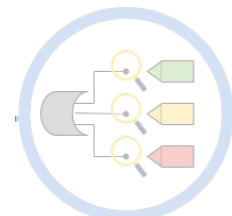
Outline



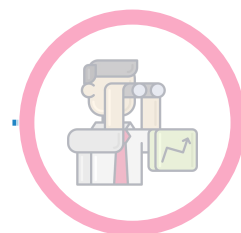
Costruzione del
Dataset



Caratterizzazione e
Modellazione del Traffico



Classificazione del
Traffico



Predizione del
Traffico



Costruzione del dataset

- Selezione delle **9 App più popolari ed utilizzate**:



Discord



GotoMeeting



Meet



Messenger



Skype



Slack



Teams



Webex



Zoom

- **3 tipologie di Attività di interesse**:
 - Audio-call (**ACall**): chiamata con solo audio tra 2 partecipanti
 - Chat (**Chat**): 2 partecipanti che si scambiano messaggi testuali e/o contenuti multimediali (es. immagini, GIF, ecc.)
 - Video-call (**VCall**): ≥ 2 partecipanti che possono trasmettere sia video che audio. Include eventi dal vivo come videochiamate, videoconferenze e webinar





Collezionamento del Traffico

- Collezionamento su **Rete GARR** (laboratorio ARCLAB dell'Università degli Studi di Napoli Federico II)
- Architettura **MIRAGE**[4]: etichettamento affidabile **biflusso-(App, Attività)**
- Partecipazione volontaria di **ricercatori** e più di **150 studenti**
- **più di 300 ore** di dati di traffico - **20 GB** di volume



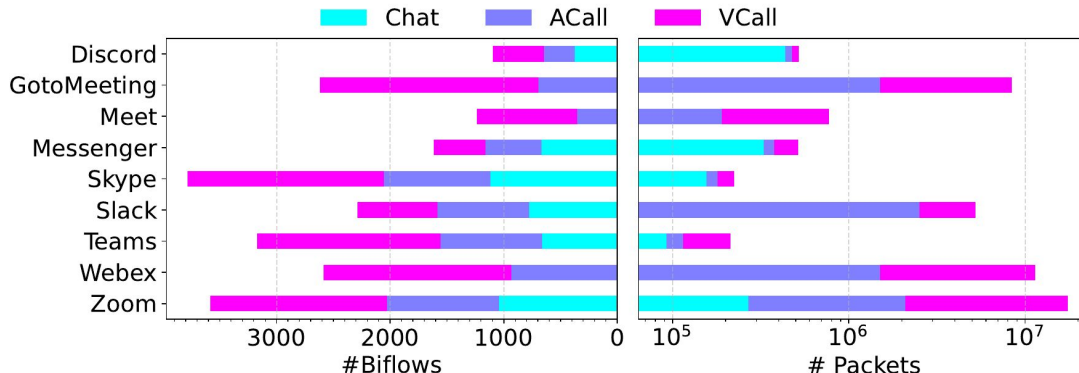
MIRAGE Project

<http://traffic.comics.unina.it/mirage/>



Privacy

Utilizzo di dispositivi e profili utente
ad-hoc





Ampliamento del Dataset

NEXT

- Incremento del **Volume**

- Nuove App e Attività



Teams



Discord



GotoMeeting



Meet



Zoom



Webex



Messenger



Skype



Slack



Whatsapp



Signal



Telegram



BlueJeans

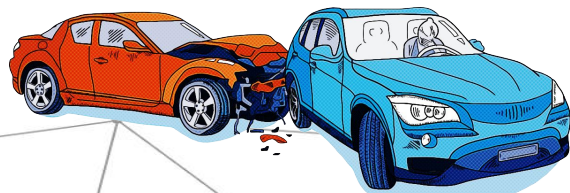


Lifesize



Jitsi Meet

- Analisi sulla **variabilità del traffico** (intervallo temporale ampio)
- Analisi sulla **robustezza** della metodologia





Outline



- Caratterizzazione a **granularità differenti** (es. **traccia**, **biflusso**)
- **Modellazione** del traffico via Markov Chains



Caratterizzazione del Traffico

- Caratterizzazione basata su **6 metriche aggregate**:
 - Upstream/Downstream bit-rate
 - Upstream/Downstream packet-rate
 - Downstream volume ratio
 - Downstream packet ratio
- ▷ Le metriche sono calcolate utilizzando **finestre temporali** (*non sovrapposte*) di durata Δ .
- ▷ Per una data traccia, la i -esima finestra considera tutti i pacchetti il cui tempo di arrivo t_k :

$$t_k \in [t_0 + (i - 1)\Delta, t_0 + i\Delta), \quad i \in \{1, 2, \dots, \lceil \frac{D}{\Delta} \rceil\}$$

dove t_0 e D sono rispettivamente il *tempo di inizio* e la *durata* della cattura.

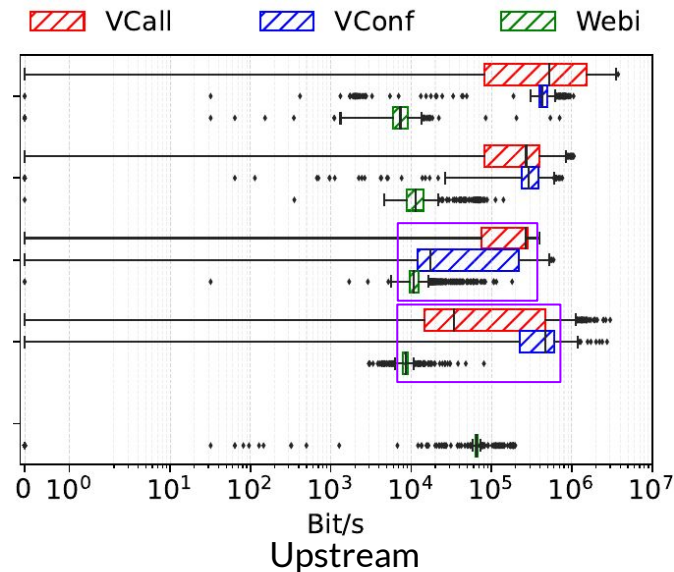
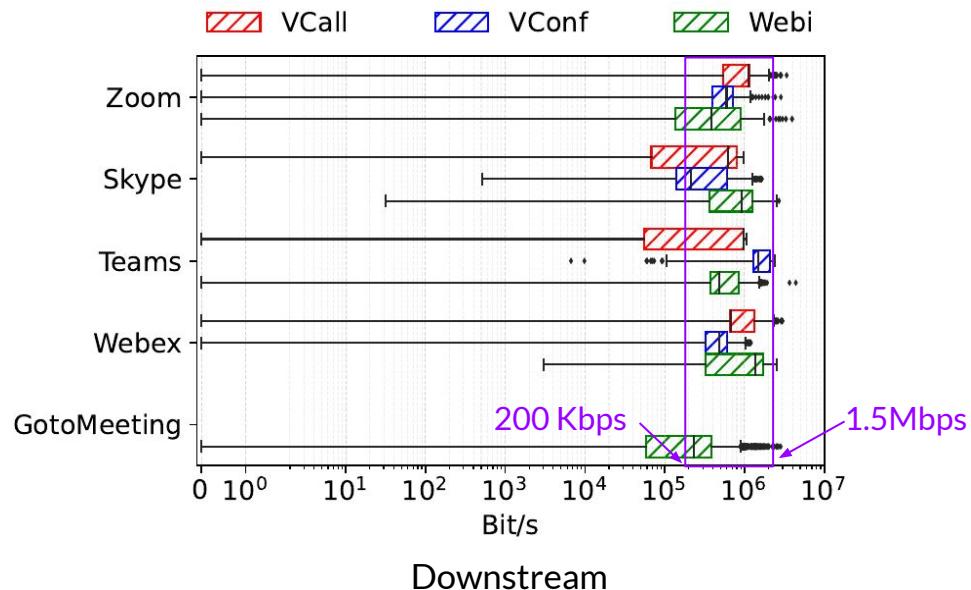
- ▷ I risultati mostrati sono ottenuti con $\Delta=5$ s (*altri valori di $\Delta= 10, 30, 60$ s sono stati testati senza mostrare differenze significative.*)



GIORNATA DI INCONTRO BORSE DI STUDIO GARR “ORIO CARLINI” BORSISTI DAY 2022



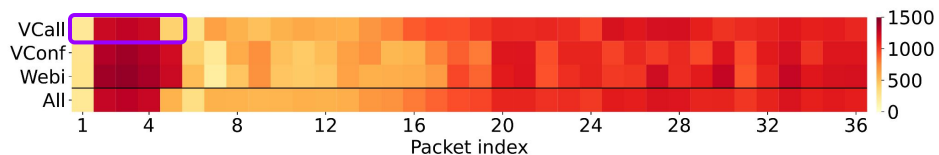
Bit-rate ($\Delta=5$ s)



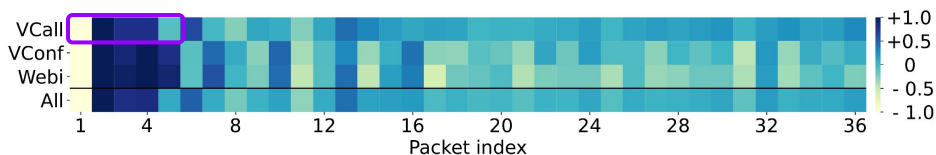
Le box riportano 1Q e 3Q, mentre i baffi 1Q-1.5IQR and 3Q+1.5IQR (IQR=3Q-1Q)



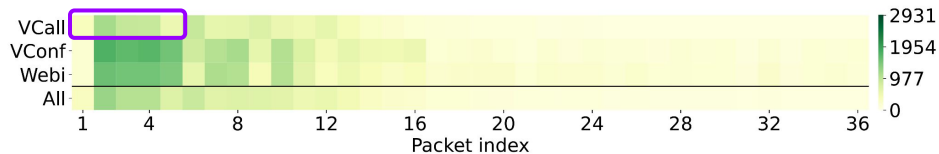
Analisi del comportamento iniziale di



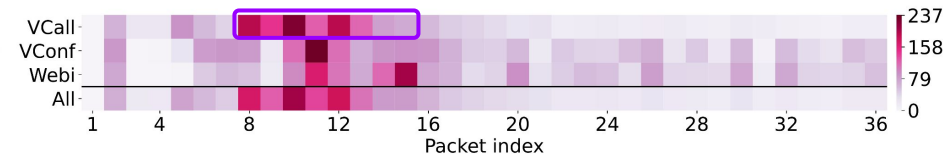
L4-Payload Length (PL) [B]



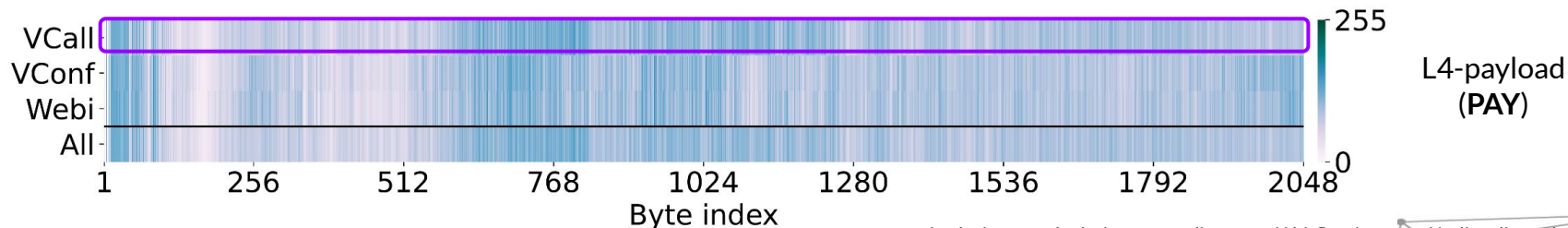
Packet Direction (DIR)



TCP Window Size (TCPWIN)



Inter-Arrival Time (IAT) [ms]



i valori sono calcolati come media su tutti i bi-flussi per ogni indice di pacchetto/byte



Feature Innovative

NEXT

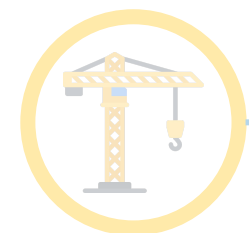


- Definizione di **feature innovative** adatte a **discriminare** in particolare le **attività** svolte dall'utente
- Analisi delle **relazioni temporali** tra i diversi biflussi





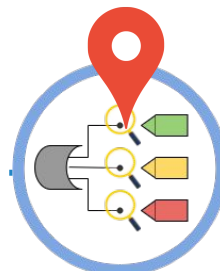
Outline



Costruzione del
Dataset

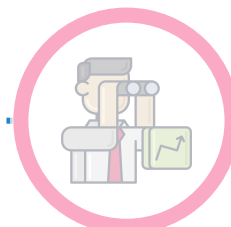


Caratterizzazione e
Modellazione del Traffico



Classificazione del
Traffico

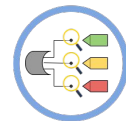
Classificazione delle
Applicazioni e delle **Attività**
svolte dall'utente



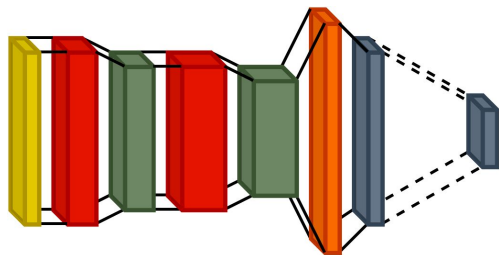
Predizione del
Traffico



Metodologia



mono-modale



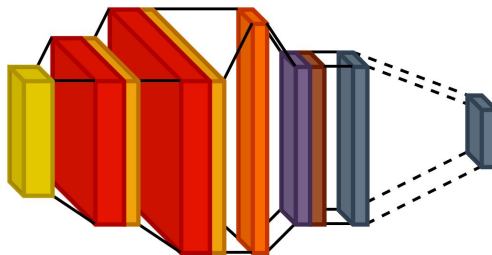
1D-CNN

[Wang et al., ISI, 2017]

Input (per ciascun biflusso):

- Primi N_b byte di PAY

mono-modale



2D-CNN+LSTM (Hybrid)

[Lopez et al, IEEE Access, 2017]

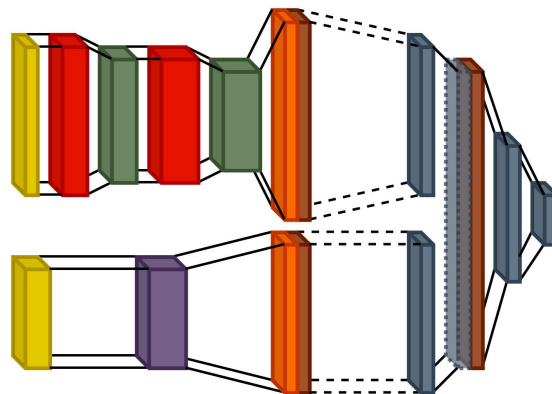
Input (per i primi N_b pacchetti di ciascun biflusso):

- **PL+IAT+DIR+TCPWIN**

Training: (loss=categorical cross-entropy, epochs=90*, batch size=50, optimizer=Adam, early-stopping=True)

*Mimetic: 2-phase training (pre-training=25 epochs per-modality, fine-tuning=40 epochs)

Multimodale



Mimetic

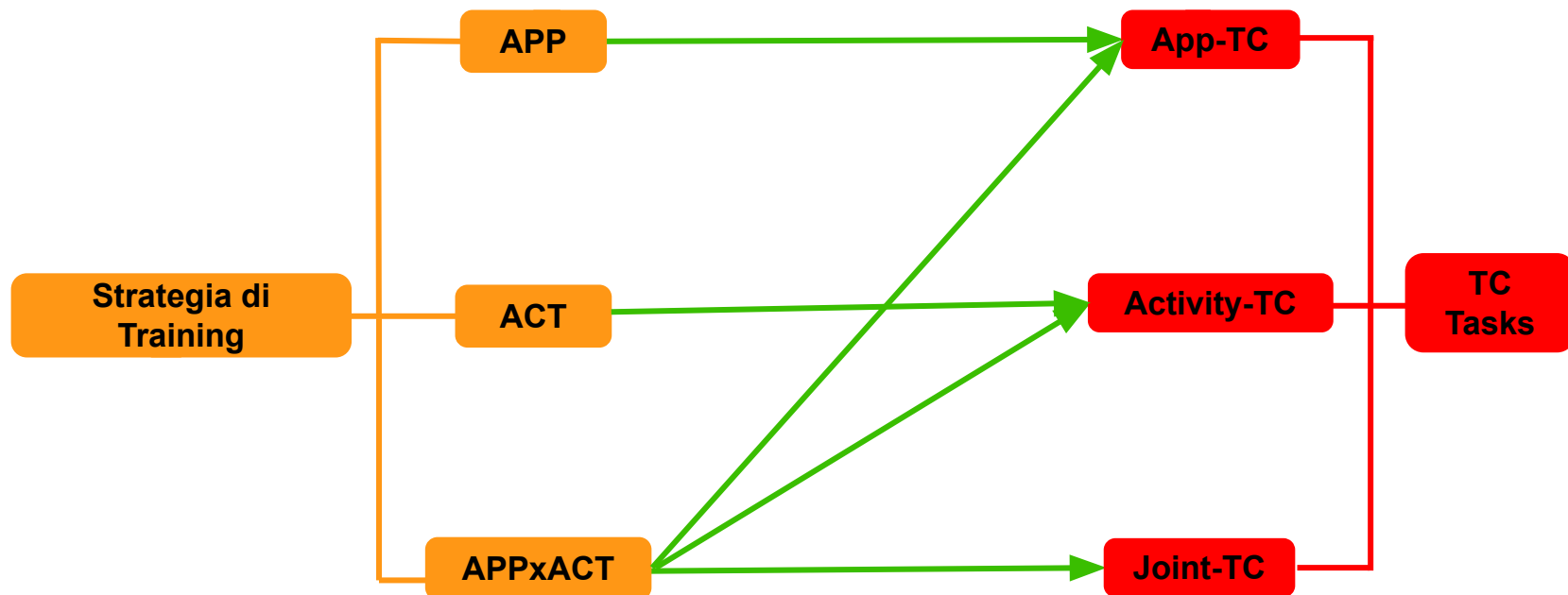
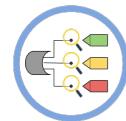
[Aceto et al., ComNet, 2019]

Input (per ciascun biflusso):

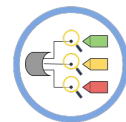
- I Modalità= 1D-CNN Input
- II Modalità= Hybrid Input



Classificazione del Traffico relativo ad **App** e **Attività**



→ in grado di risolvere



Risultati Sperimentali



GotoMeeting



Skype



Teams



Webex



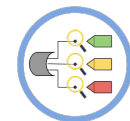
Zoom

Classifier	N_b	N_p	Strategy	Joint-TC	App-TC	Act-TC	#TP[k]
				F-Measure [%]	F-Measure [%]	F-Measure [%]	
1D-CNN	576	-	APPxACT	49.74	96.98	66.62	4261
			APP		97.89		4253
			ACT			65.89	4251
Hybrid	-	12	APPxACT	49.99	95.08	63.38	222
			APP		94.71		222
			ACT			62.50	221
MIMETIC	576	12	APPxACT	52.71	97.86	67.48	942
			APP		98.30		937
			ACT			66.87	936

Attività:

- Video-Call (VCall)
- Video-Conference (VConf)
- Webinar (Webi)

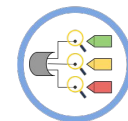
I risultati sono riportati
come media ottenuta su 10
fold



Risultati Sperimentali

Classifier	N_b	N_p	Strategy	Joint-TC	App-TC	Act-TC	#TP[k]
				F-Measure [%]	F-Measure [%]	F-Measure [%]	
1D-CNN	576	-	APPxACT	49.74	96.98	66.62	4261
			APP		97.89		4253
			ACT			65.89	4251
Hybrid	-	12	APPxACT	49.99	95.08	63.38	222
			APP		94.71		222
			ACT			62.50	221
MIMETIC	576	12	APPxACT	52.71	97.86	67.48	942
			APP		98.30		937
			ACT			66.87	936

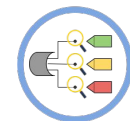
I risultati sono riportati
come media ottenuta su 10
fold



Risultati Sperimentali

Classifier	N_b	N_p	Strategy	Joint-TC	App-TC	Act-TC	#TP[k]
				F-Measure [%]	F-Measure [%]	F-Measure [%]	
1D-CNN	576	-	APPxACT	49.74	96.98	66.62	4261
			APP		97.89		4253
			ACT			65.89	4251
Hybrid	-	12	APPxACT	49.99	95.08	63.38	222
			APP		94.71		222
			ACT			62.50	221
MIMETIC	576	12	APPxACT	52.71	97.86	67.48	942
			APP		98.30		937
			ACT			66.87	936

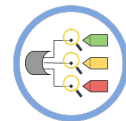
I risultati sono riportati
come media ottenuta su 10
fold



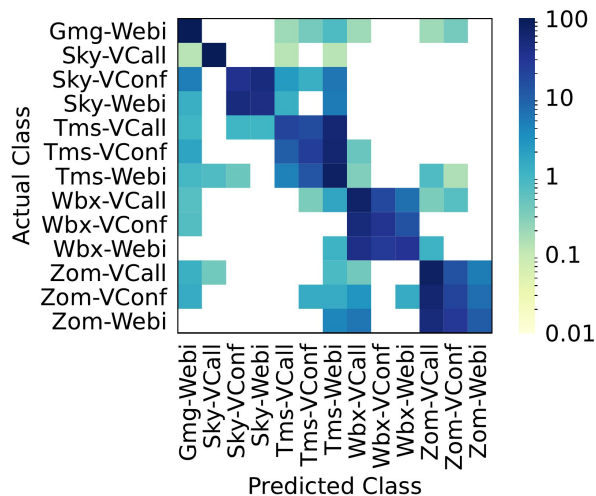
Risultati Sperimentali

Classifier	N_b	N_p	Strategy	Joint-TC	App-TC	Act-TC	#TP[k]
				F-Measure [%]	F-Measure [%]	F-Measure [%]	
1D-CNN	576	-	APPxACT	49.74	96.98	66.62	4261
			APP		97.89		4253
			ACT			65.89	4251
Hybrid	-	12	APPxACT	49.99	95.08	63.38	222
			APP		94.71		222
			ACT			62.50	221
	576	12	APPxACT	52.71	97.86	67.48	942
			APP		98.30		937
			ACT			66.87	936

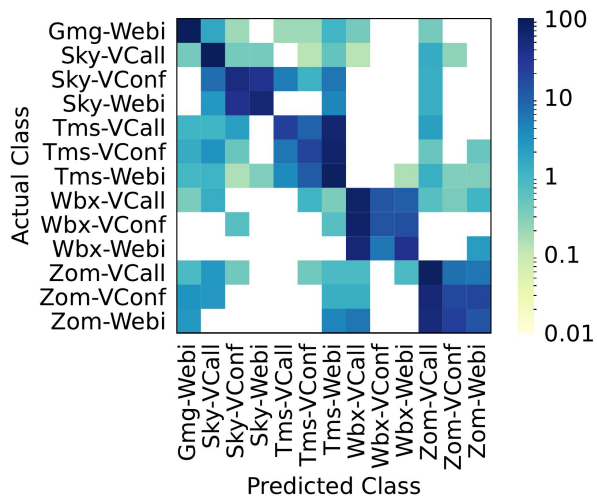
I risultati sono riportati
come media ottenuta su 10
fold



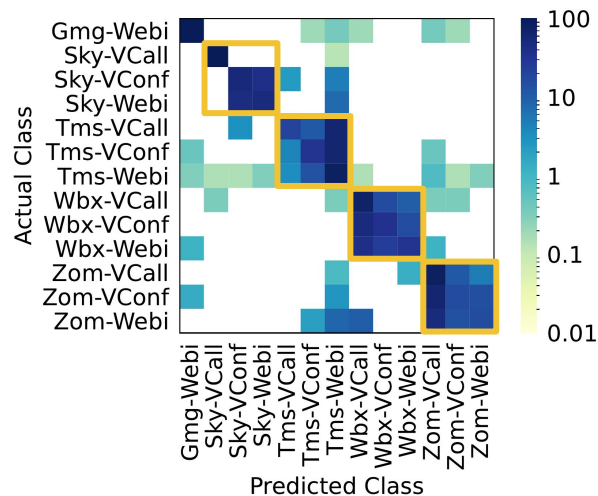
Errore di Misclassificazione (Joint-TC)



1D-CNN



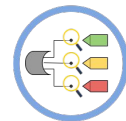
Hybrid



MIMETIC



I risultati sono riportati come media ottenuta su 10 fold



Miglioramento della Metodologia di Classificazione

NEXT

- Inclusione delle **nuove feature**
- Utilizzo di **livelli avanzati** di Deep-Learning (es. embedding, attention ecc.)



- **Ri-progettazione** delle architetture (adattamento o estensione)
- **Validazione** delle architetture (trade-off prestazioni - complessità)



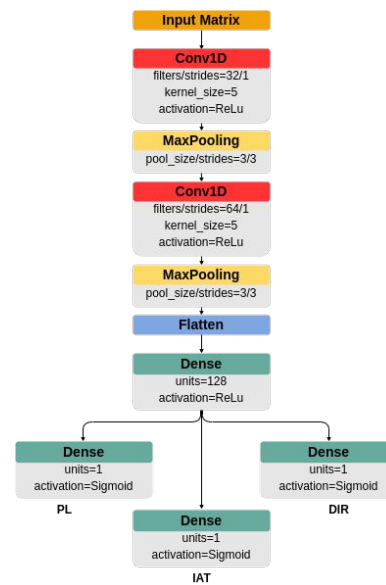
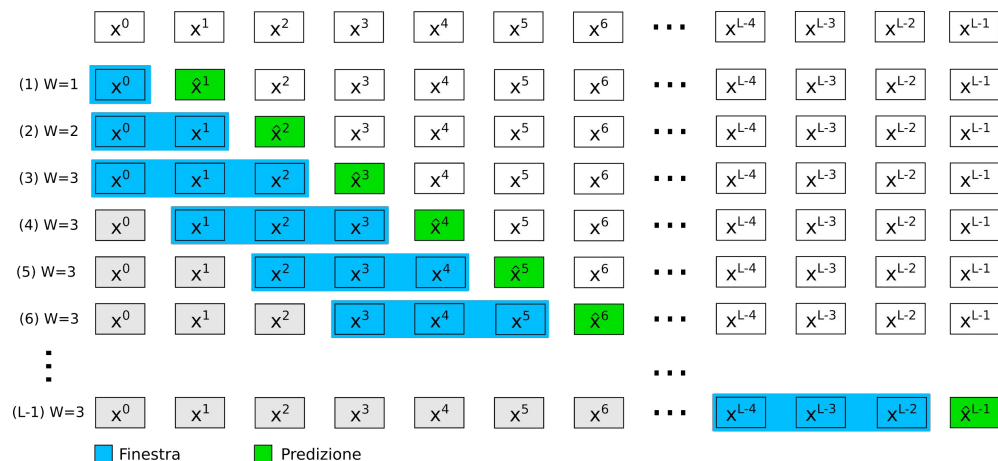


Outline





Metodologia di Predizione



Input

- Sequenza di PL, DIR e IAT

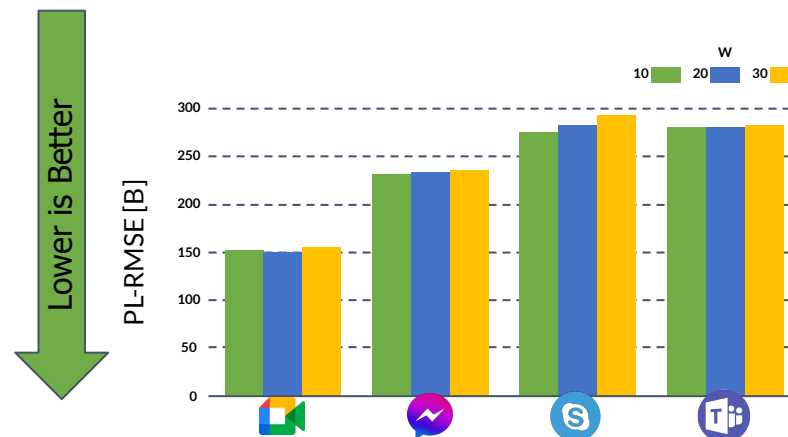
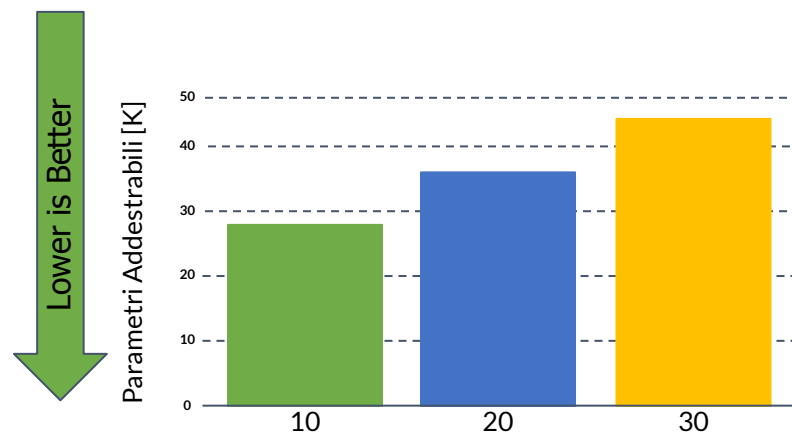
Output

- Valore predetto di PL, DIR e IAT per il pacchetto $W+1$

Training: (loss=mean_squared_error (PL/IAT) - binary_crossentropy (DIR), epochs=100, batch size=50, optimizer=Adam, early-stopping=True)



Prestazioni al variare della finestra





Miglioramento della Metodologia di Predizione

NEXT

- Progetto e Validazione di modelli in grado di **sfruttare** la **conoscenza** riguardo l'**attività** svolta dall'utente
- Progetto e Validazione di **modelli a granularità differenti** (es. combinazione App-Attività)
- Utilizzo di **nuove feature** e **tecniche multimodali**





Conclusioni e Sviluppi Futuri

Caratterizzazione e Modellazione del traffico

- **Take home message:** *le attività associate alla medesima App non sono facilmente distinguibili*
- **ToDo:** definizione di feature innovative

Classificazione del traffico associato ad App e Attività

- **Take home message:**
 - *buone prestazioni relative alla classificazione delle App (F-Measure $\geq 95\%$)*
 - *la classificazione delle attività richiede ulteriori ricerche*
- **ToDo:** miglioramento della metodologia di Classificazione (es. nuove feature , livelli avanzati di Deep Learning ecc.)



Conclusioni e Sviluppi Futuri

Predizione del traffico (a grana fine)

- **Take home message:** *Le prestazioni variano a seconda dell'App considerata e della feature predetta*
- **ToDo:** miglioramento della metodologia di Predizione (es. modelli a granularità differente, feature aggiuntive ecc.)

Ampliamento del Dataset

- Ampliamento del dataset (es. nuove App e Attività)
- Analisi sulla **variazione del traffico** e sulla **robustezza** dei modelli (concept drift)
- Rilascio pubblico



GIORNATA DI INCONTRO BORSE DI STUDIO GARR “ORIO CARLINI” BORSISTI DAY 2022

Grazie per
l'attenzione!

Domande?

