



Corso di Laurea in Informatica

Single e Cross-layer Detection di Siti Web Malevoli: Un Confronto Empirico

Prof. Fabio Palomba

Nicolapio Gagliarde
Mat.: 0512106980



Il problema



170 milioni di siti web
malevoli



Il problema



170 milioni di siti web
malevoli



650 milioni di attacchi



Il problema



170 milioni di siti web
malevoli



650 milioni di attacchi



Milioni di dollari persi al minuto





Analisi degli URL



Analisi degli URL



Analisi dei redirect e delle risorse richieste



n.gagliarde@studenti.unisa.it



<https://github.com/GagliardeNicolapio>



<https://www.linkedin.com/in/nicolapio-gagliarde-75209018b/>



Analisi degli URL



Analisi dei redirect e delle risorse richieste



Analisi della pagina web



Analisi degli URL



Analisi dei redirect e delle risorse richieste



Analisi della pagina web

Tecniche
Single-layer

Potrebbero fallire con:



Analisi degli URL



Indirizzi corti o troppo simili a URL benevoli



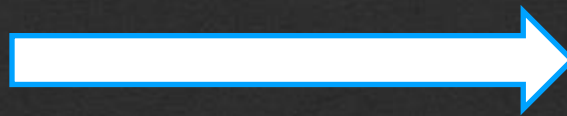
Analisi dei redirect e delle risorse richieste



Analisi della pagina web



Analisi degli URL



Potrebbero fallire con:

Indirizzi corti o troppo simili a URL benevoli



Analisi dei redirect e delle risorse richieste



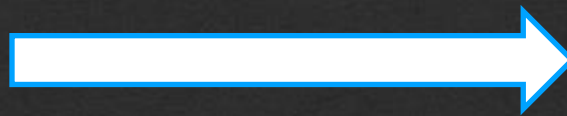
Siti creati con i CMS



Analisi della pagina web



Analisi degli URL



Potrebbero fallire con:

Indirizzi corti o troppo simili a URL benevoli



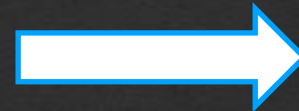
Analisi dei redirect e delle risorse richieste



Siti creati con i CMS




Analisi della pagina web



Siti creati con un interfaccia clonata e tecniche di offuscamento

Single e Cross-layer



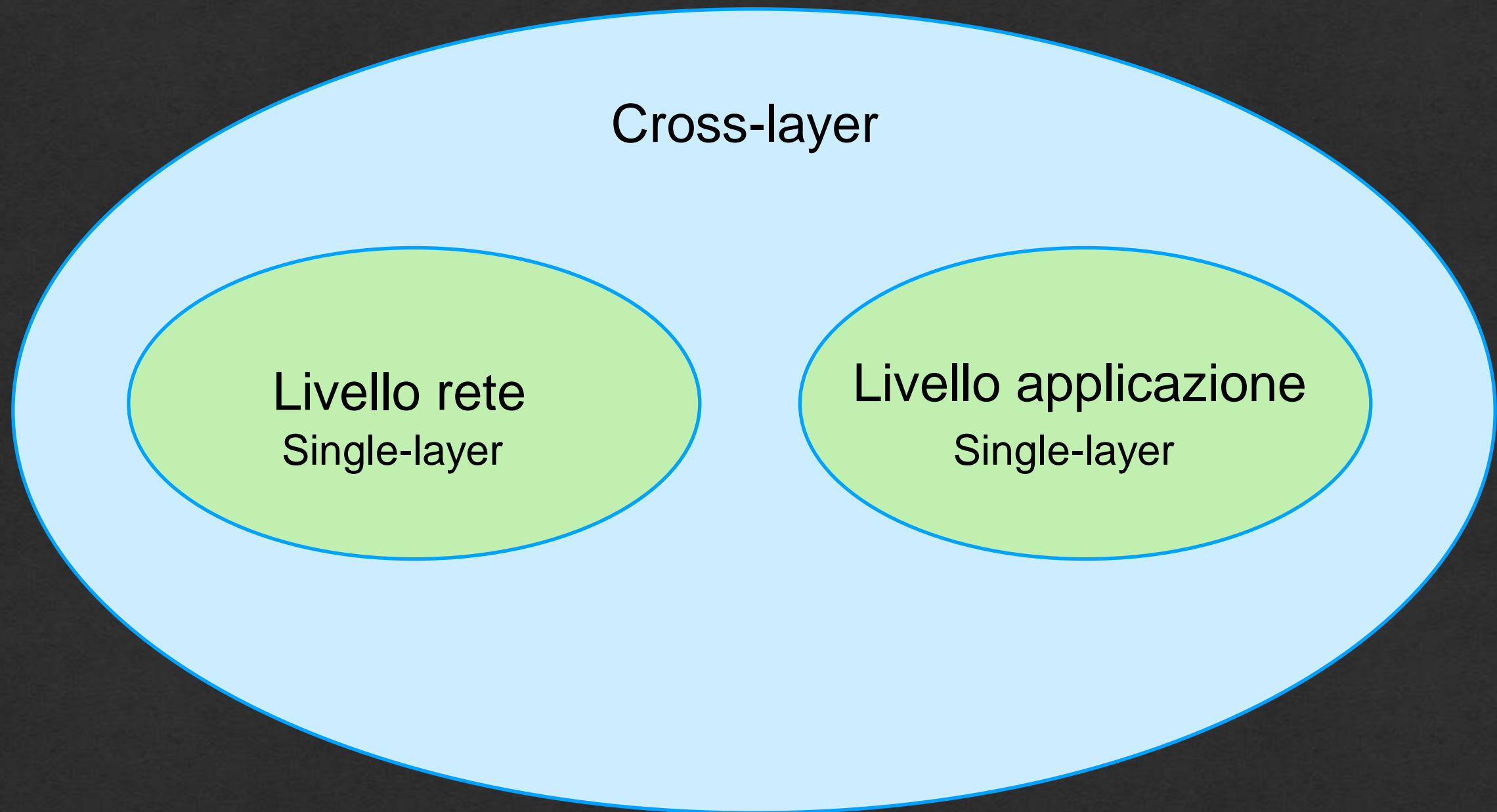
Livello rete
Single-layer

Single e Cross-layer

Livello rete
Single-layer

Livello applicazione
Single-layer

Single e Cross-layer



Lo scopo e le differenze

Lo scopo: confrontare i risultati
ottenuti con i risultati di
Xu[2014]¹



¹ Li Xu. Detecting and characterizing malicious websites.
The University of Texas at San Antonio, 2014.

Lo scopo e le differenze

Lo scopo: confrontare i risultati ottenuti con i risultati di Xu[2014]¹

- medesimi algoritmi

¹ Li Xu. Detecting and characterizing malicious websites. The University of Texas at San Antonio, 2014.



Lo scopo e le differenze

Lo scopo: confrontare i risultati ottenuti con i risultati di Xu[2014]¹

- medesimi algoritmi
- medesime tecniche

¹ Li Xu. Detecting and characterizing malicious websites. The University of Texas at San Antonio, 2014.



Lo scopo e le differenze

Lo scopo: confrontare i risultati ottenuti con i risultati di Xu[2014]¹

- medesimi algoritmi
- medesime tecniche
- dataset diverso!

¹ Li Xu. Detecting and characterizing malicious websites. The University of Texas at San Antonio, 2014.

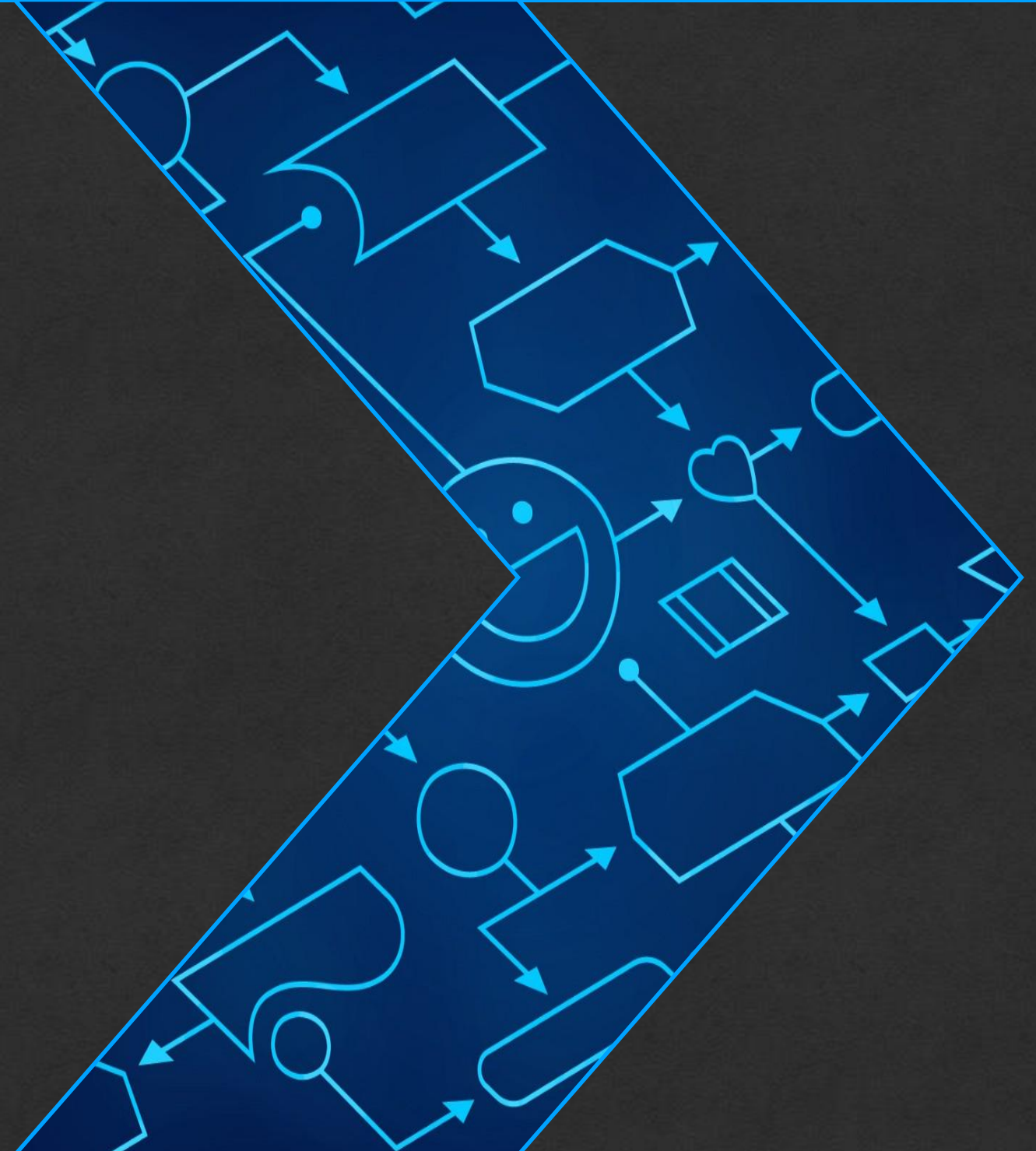


Naive Bayes

Logistic Regression

Support Vector Machine

Decision Tree



Principal Component Analysis

CFS Subset Evaluation

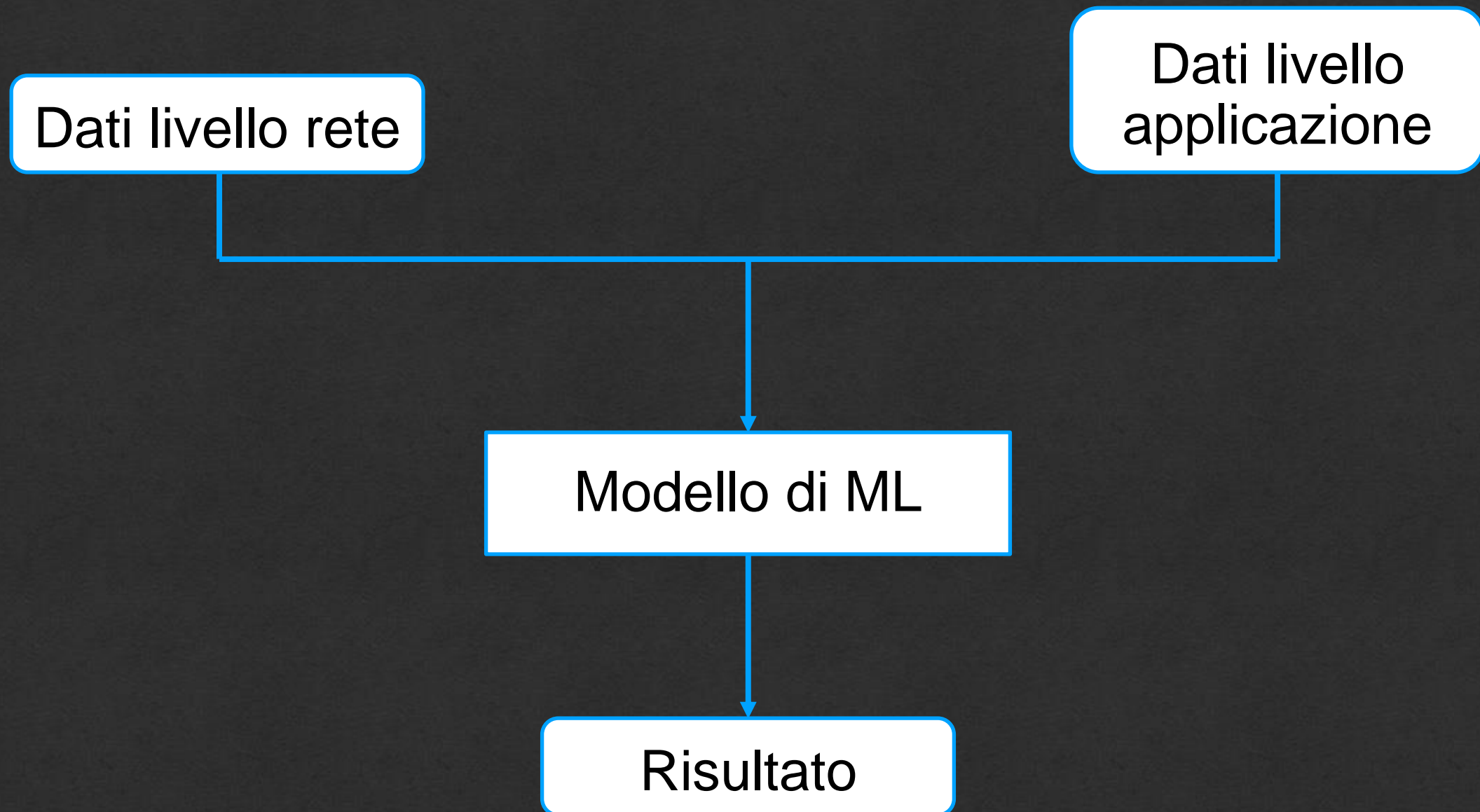
Information Gain

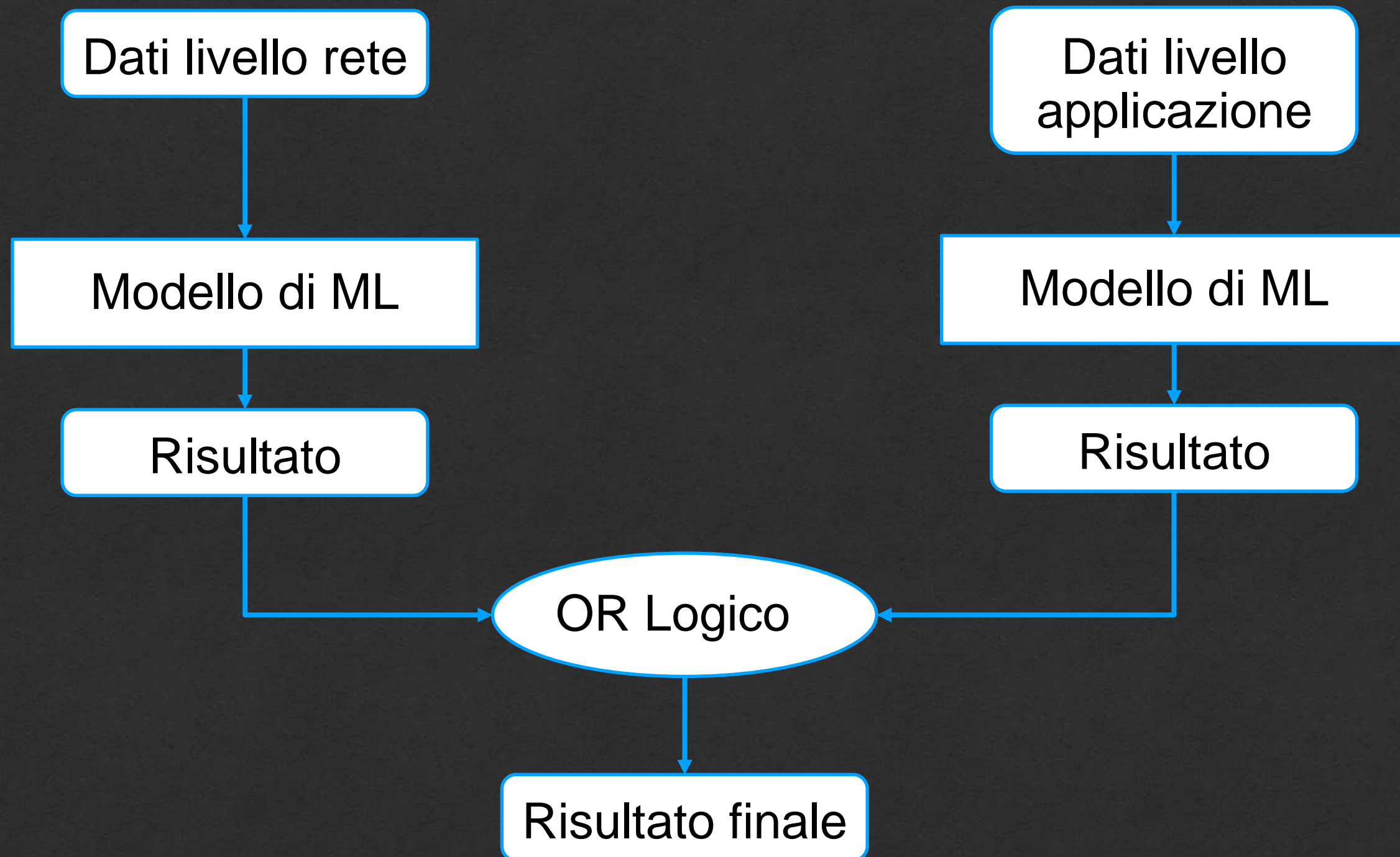
Per il cross-layer

Data-aggregation

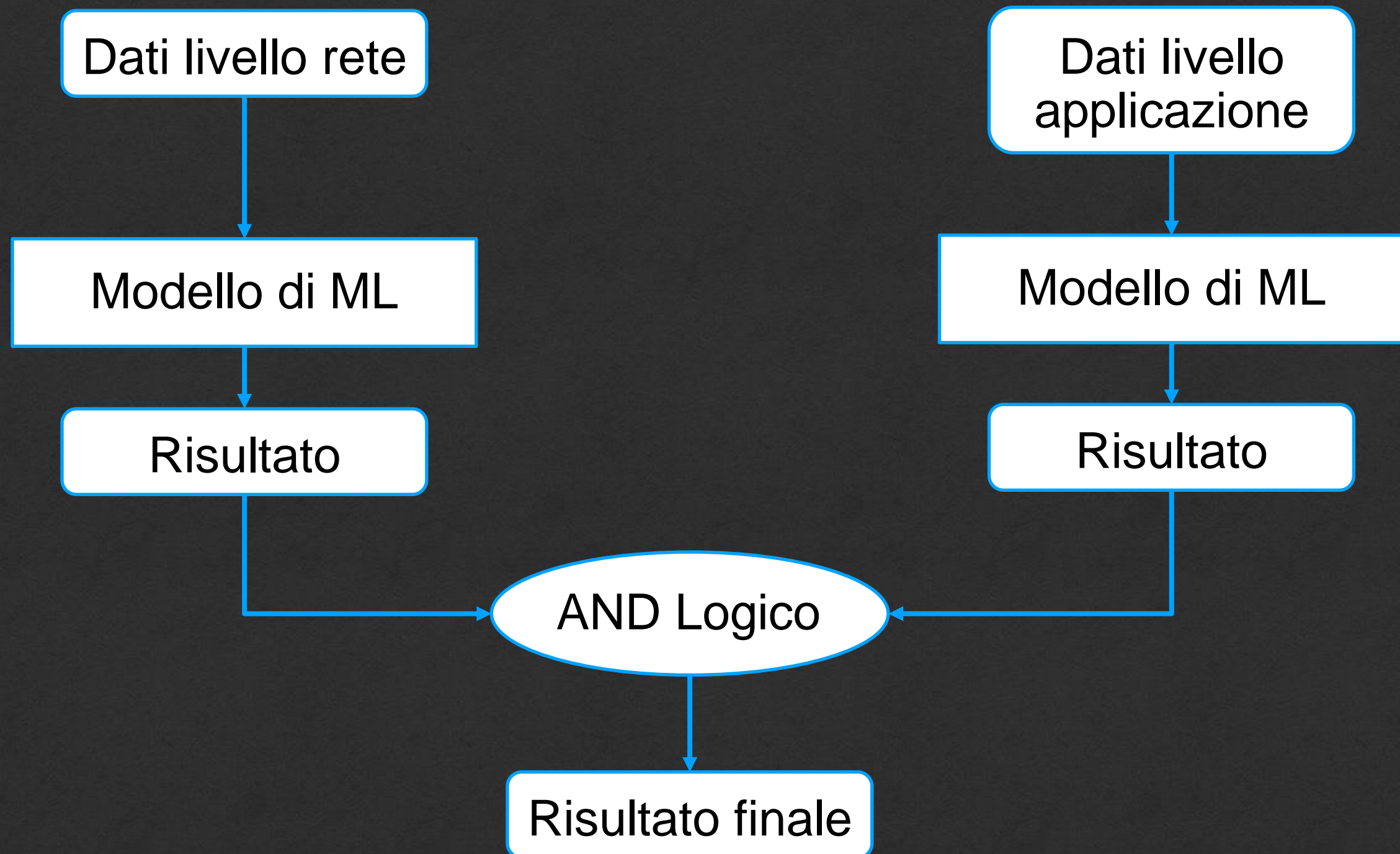
OR-aggregation

AND-aggregation





AND-aggregation



Algoritmo migliore: Decision Tree

- Accuracy: 95%
- Falsi negativi: 5%
- Falsi positivi: 7%



Algoritmo migliore: Decision Tree

La AND-aggregation con Decision Tree implica un incremento dei falsi negativi di circa 20 punti percentuali

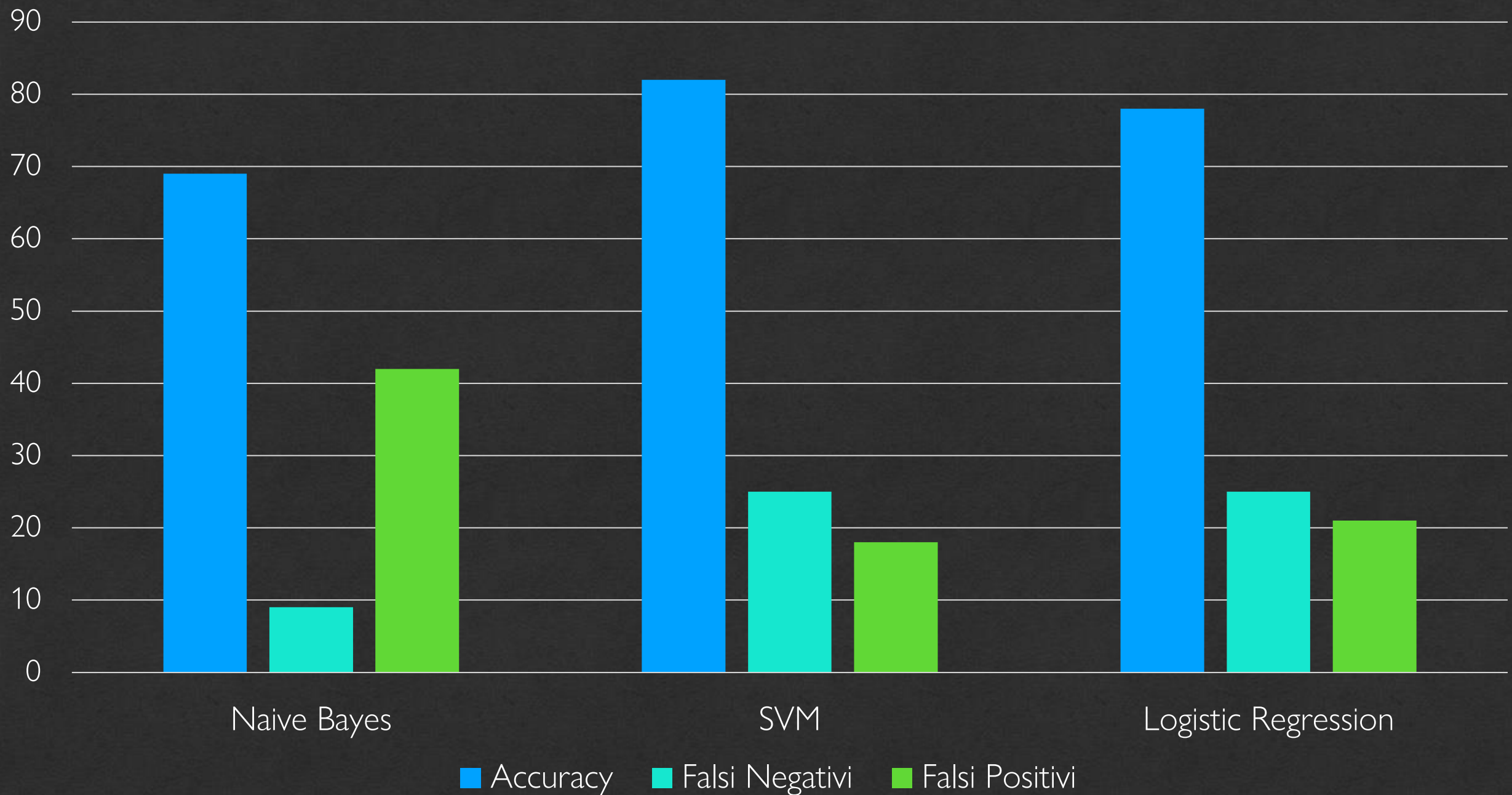


Algoritmo migliore: Decision Tree

La AND-aggregation con Decision Tree implica un incremento dei falsi negativi di circa 20 punti percentuali

Naive Bayes, SVM e Logistic Regression risultano non adatti alla classificazione single e cross-layer





Il problema



- 170 milioni di siti web malevoli
- 650 milioni di attacchi
- Milioni di dollari persi al minuto

Single e Cross-layer Detection di Siti Web Malevoli: Un Confronto Empirico
Nicolapio Gagliarde
Università degli Studi di Salerno

Alcune soluzioni



- Analisi degli URL
- Analisi dei redirect e delle risorse richieste
- Analisi della pagina web

Tecniche Single-layer

Single e Cross-layer Detection di Siti Web Malevoli: Un Confronto Empirico
Nicolapio Gagliarde
Università degli Studi di Salerno

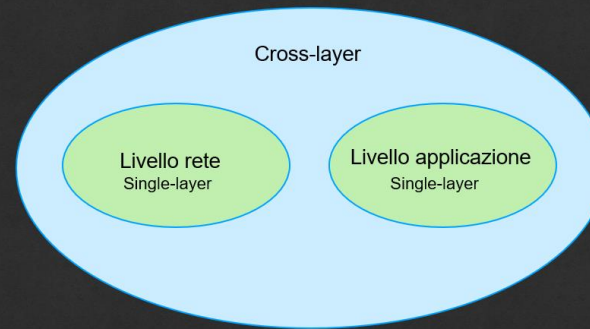
Alcune soluzioni



- Potrebbero fallire con:
- Analisi degli URL → Indirizzi corti o troppo simili a URL benevoli
 - Analisi dei redirect e delle risorse richieste → Siti creati con i CMS
 - Analisi della pagina web → Siti creati con un'interfaccia clonata e tecniche di offuscamento

Single e Cross-layer Detection di Siti Web Malevoli: Un Confronto Empirico
Nicolapio Gagliarde
Università degli Studi di Salerno

Single e Cross-layer



Single e Cross-layer Detection di Siti Web Malevoli: Un Confronto Empirico
Nicolapio Gagliarde
Università degli Studi di Salerno

Lo scopo e le differenze



Lo scopo: confrontare i risultati ottenuti con i risultati di Xu[2014]¹

- medesimi algoritmi
- medesime tecniche
- dataset diverso!

¹ Li Xu, Detecting and characterizing malicious websites. The University of Texas at San Antonio, 2014.

Single e Cross-layer Detection di Siti Web Malevoli: Un Confronto Empirico
Nicolapio Gagliarde
Università degli Studi di Salerno

I risultati



Algoritmo migliore: Decision Tree

La AND-aggregation con Decision Tree implica un incremento dei falsi negativi di circa 20 punti percentuali

Naive Bayes, SVM e Logistic Regression risultano non adatti alla classificazione single e cross-layer

Single e Cross-layer Detection di Siti Web Malevoli: Un Confronto Empirico
Nicolapio Gagliarde
Università degli Studi di Salerno

Single e Cross-layer Detection di Siti Web Malevoli: Un Confronto Empirico

Grazie!

Nicolapio Gagliarde

n.gagliarde@studenti.unisa.it



<https://github.com/GagliardeNicolapio>



[https://www.linkedin.com/in/nicolapio-](https://www.linkedin.com/in/nicolapio-gagliarde-75209018b/)



[gagliarde-75209018b/](https://www.linkedin.com/in/nicolapio-gagliarde-75209018b/)