



Corso di Laurea (Magistrale) in Informatica

Estrazione di requisiti non funzionali di sicurezza tramite tecniche di NLP

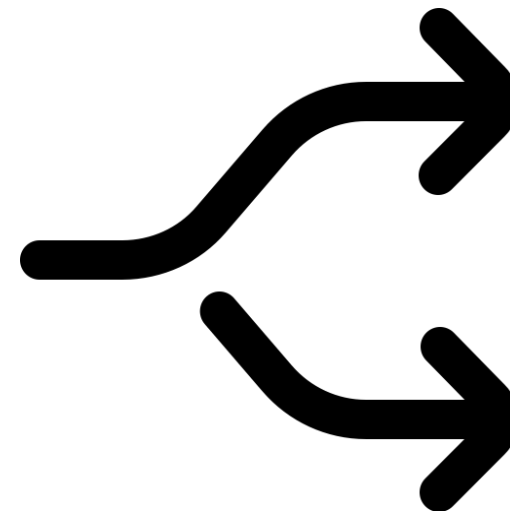
Prof. Fabio Palomba
Dott. Francesco Casillo

Dario Mazza
Mat.: 0512106742





**INGEGNERIA
DEI REQUISITI**



**REQUISITI
FUNZIONALI**

**REQUISITI NON
FUNZIONALI**



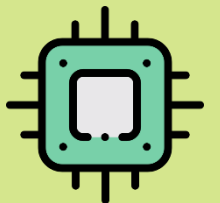
REQUISITI NON FUNZIONALI

REQUISITI NON FUNZIONALI



Rappresentano i vincoli e le proprietà relative al sistema...

...come il numero di processi che il sistema può gestire, quali sono i problemi di sicurezza di cui il sistema deve occuparsi ecc.



Sono spesso più critici dei requisiti funzionali

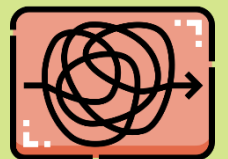
REQUISITI DI SICUREZZA

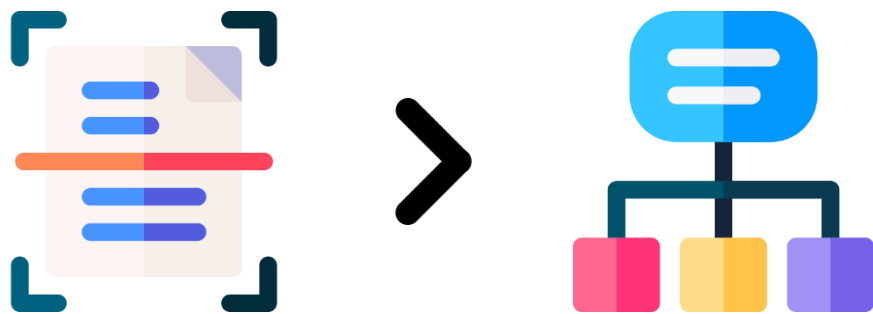
Determinano la robustezza e l'affidabilità del software



Tendono ad essere considerati in fasi avanzate del CVS compromettendo la sicurezza del sistema

Spesso distribuiti in tutta la documentazione relativa ai requisiti

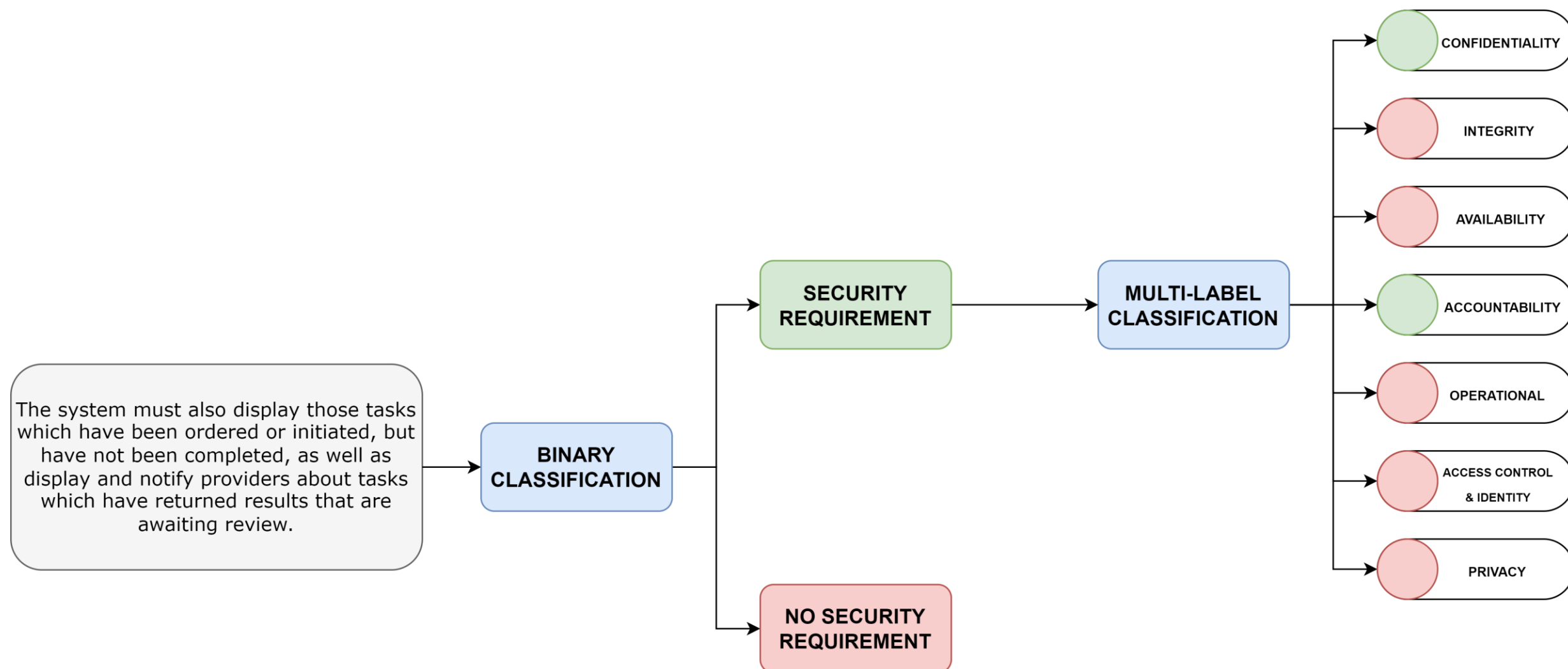




Uno **strumento** per l'identificazione
classificazione di categorie dei
requisiti di sicurezza.

In modo da aiutare gli **ingegneri del software** nella fase di definizione dei requisiti.







Dataset di Riaz *et al.*



“Hidden in Plain Sight: Automatically Identifying Security Requirements from Natural Language Artifacts”

10963 Istanze

10 Categorie

675 Istanze duplicate

3352 Istanze che non sono requisiti





3352 Istanze che non sono requisiti



Prototipi di funzioni
o metodi, errori e
messaggi log



Titoli paragrafi e
didascalie tabelle o
immagini



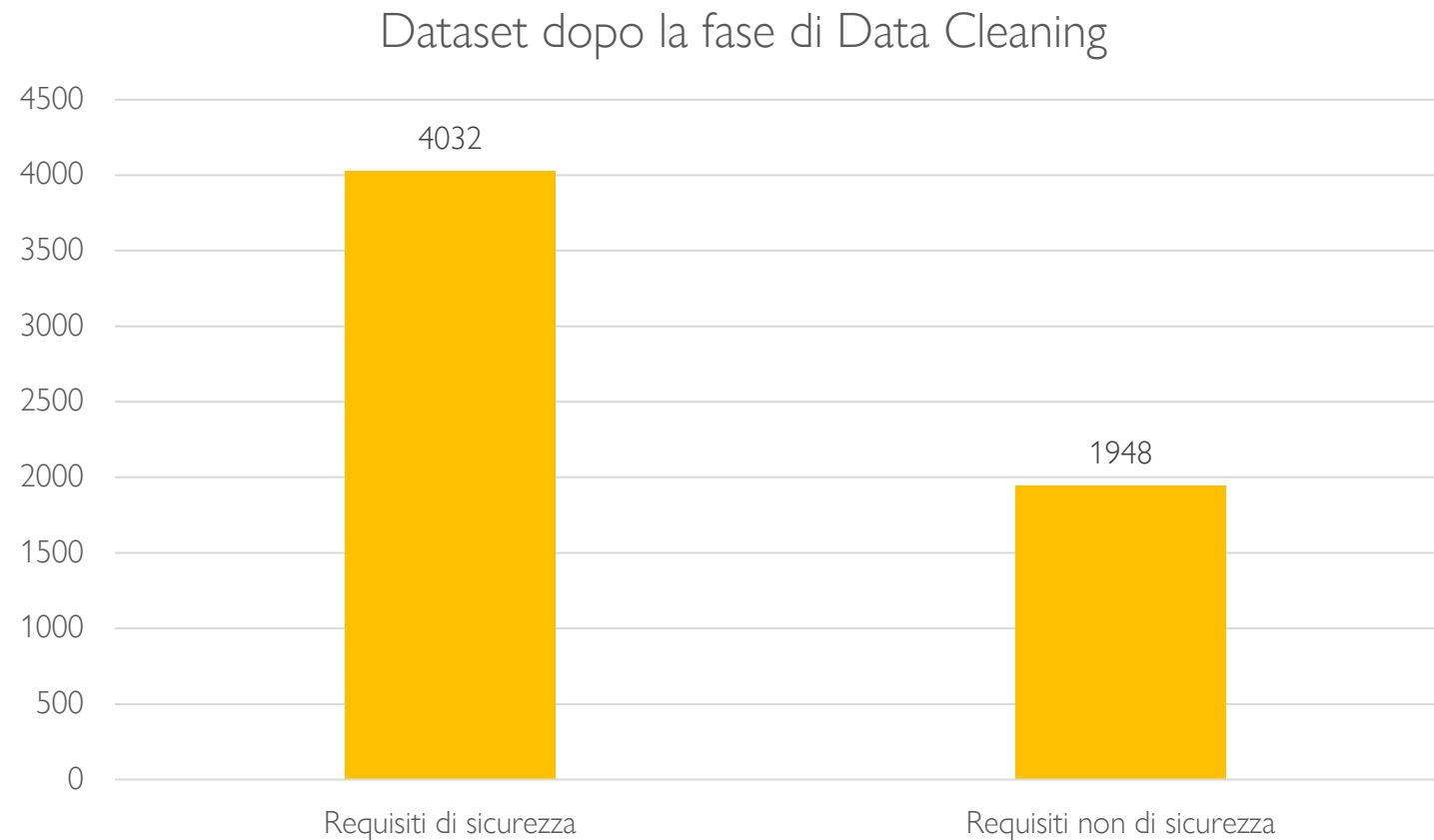
Informazioni
temporali come
date e orari



Solo caratteri
speciali, email
e url

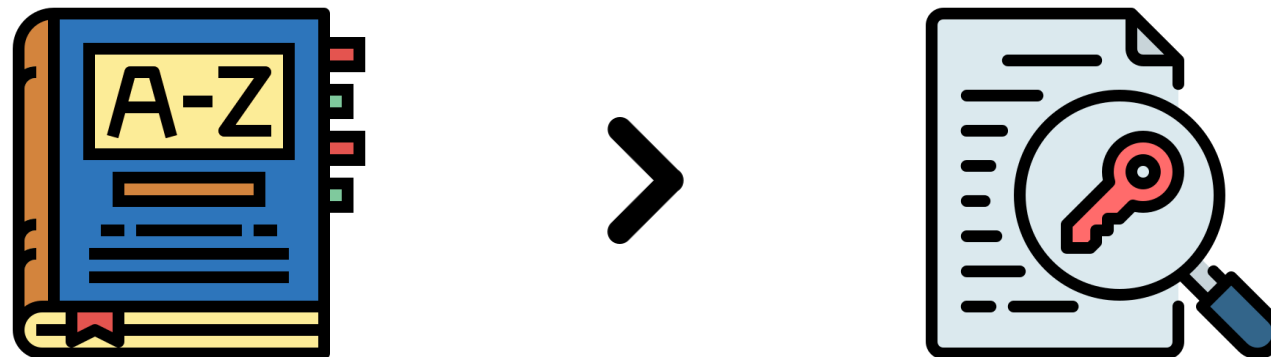


5980 Requisiti



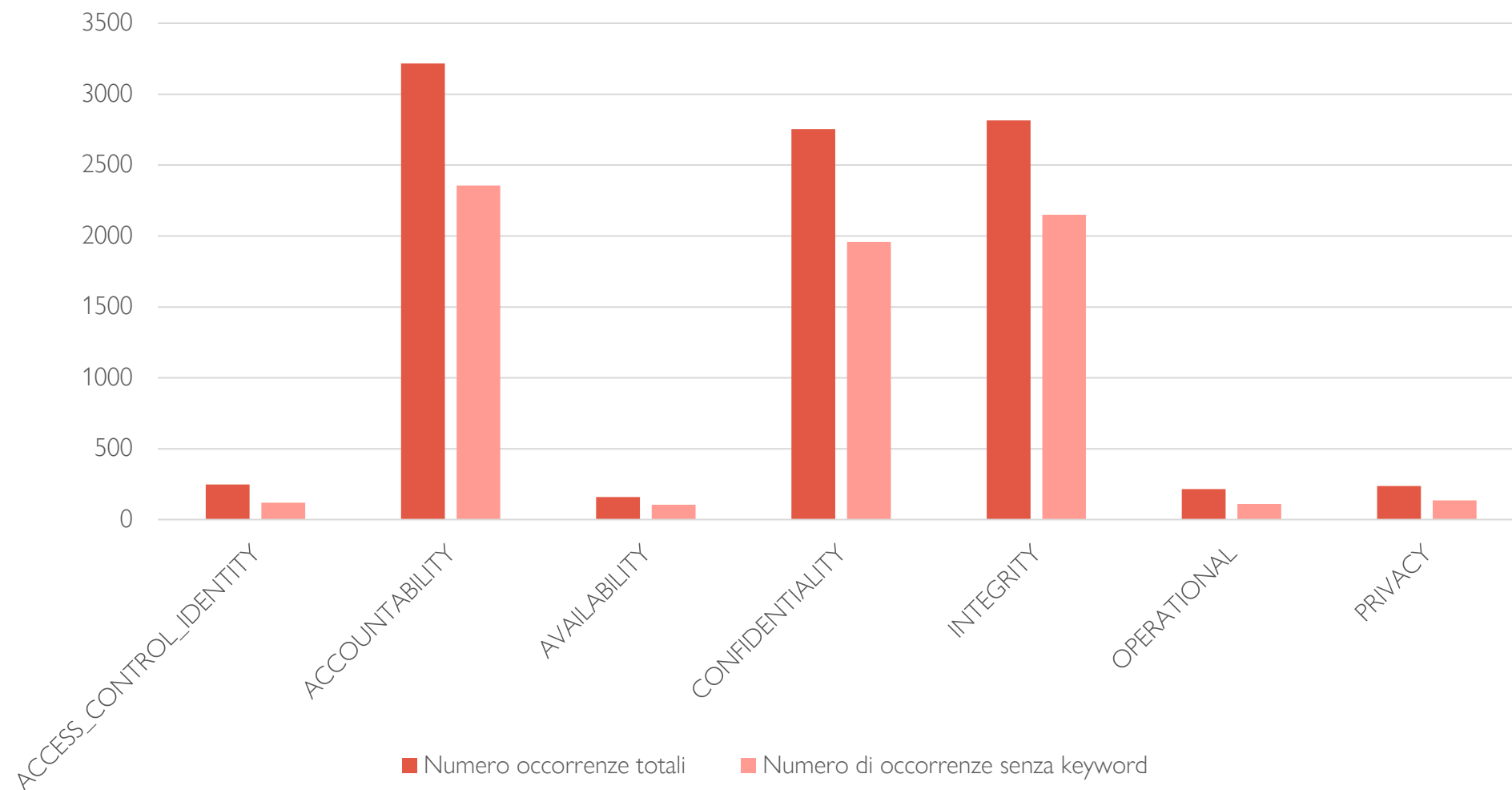


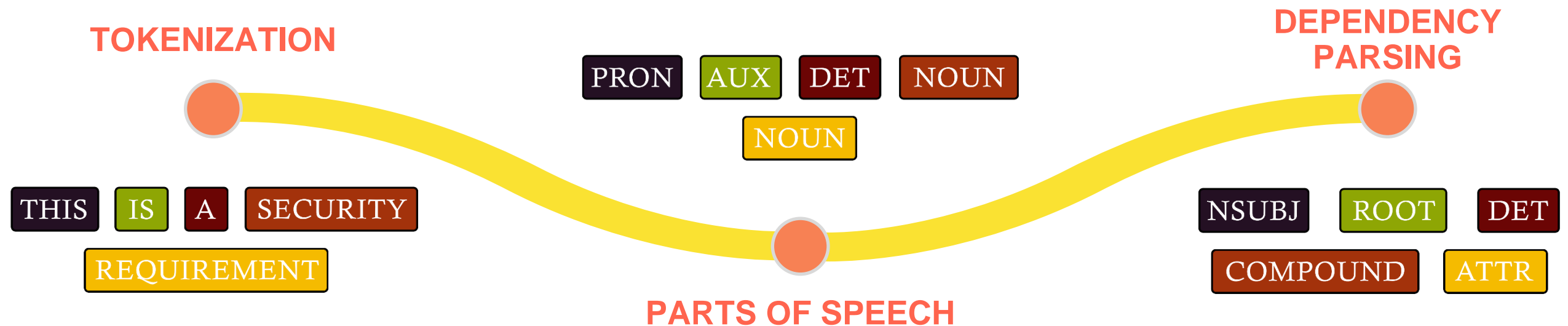
Dizionario security keywords





Analisi statistica security keywords





Scelta delle categorie

Metodologia



Confidentiality



Accountability



Integrity



Operational



Availability



**Access Control
& Authentication**

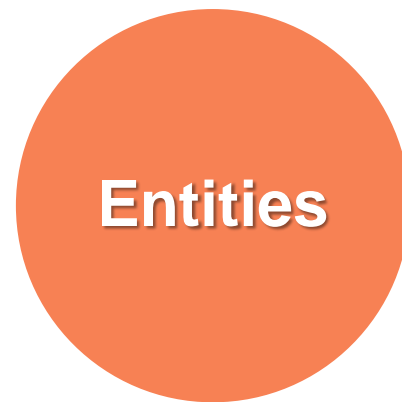
PRIVATE



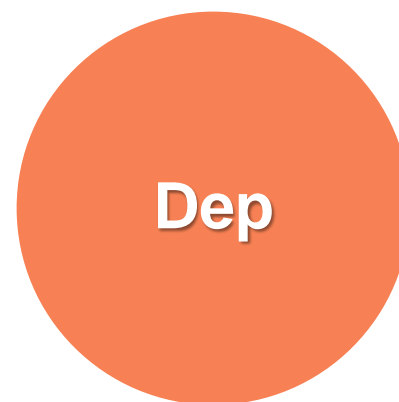
Privacy

Selezione e codifica delle features

Metodologia



Variabile
Dipendente



Variabile
Dipendente



Variabile
Dipendente



Variabile
Indipendente



Variabile
Indipendente



Codifica delle features numeri interi tramite Tokenizer

Buone prestazioni a prescindere
dal dominio applicativo

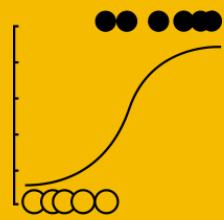


Vettorizzazione e trasformazione delle features tramite *TfidfVectorizer*

Ottime prestazioni ma
fortemente dipendente dal
dominio applicativo su cui i
modelli sono stati allenati



Random Forest



Logistic Regression



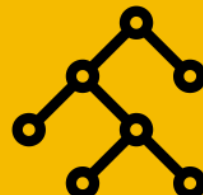
K-Nearest Neighbors



LinearSVC



Gaussian Naive-Bayes



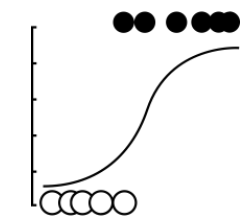
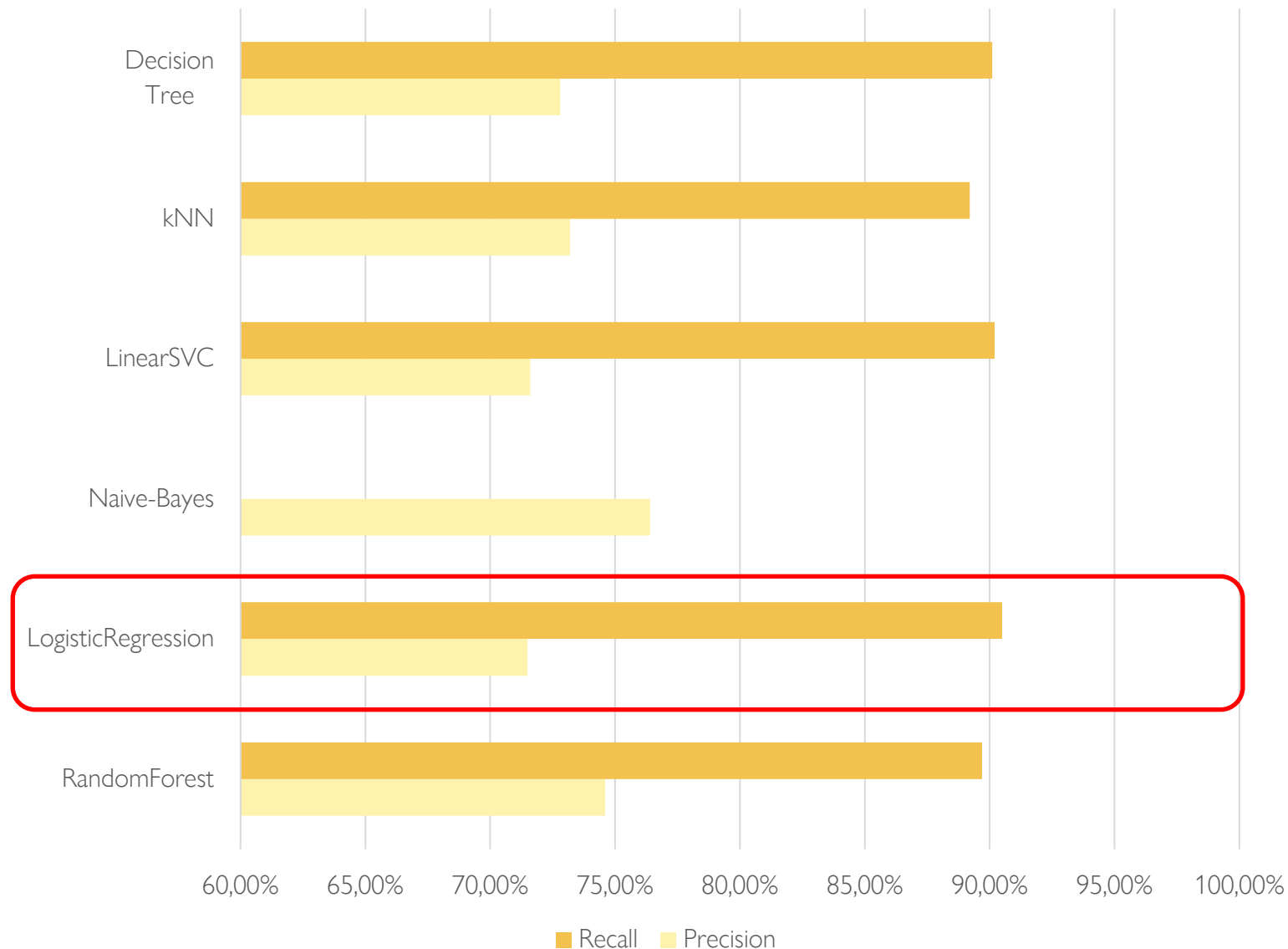
Decision Tree

Binary Relevance

Classifier Chain

Label Powerset

Risultati con Tokenizer



Logistic
Regression

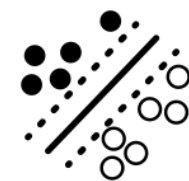
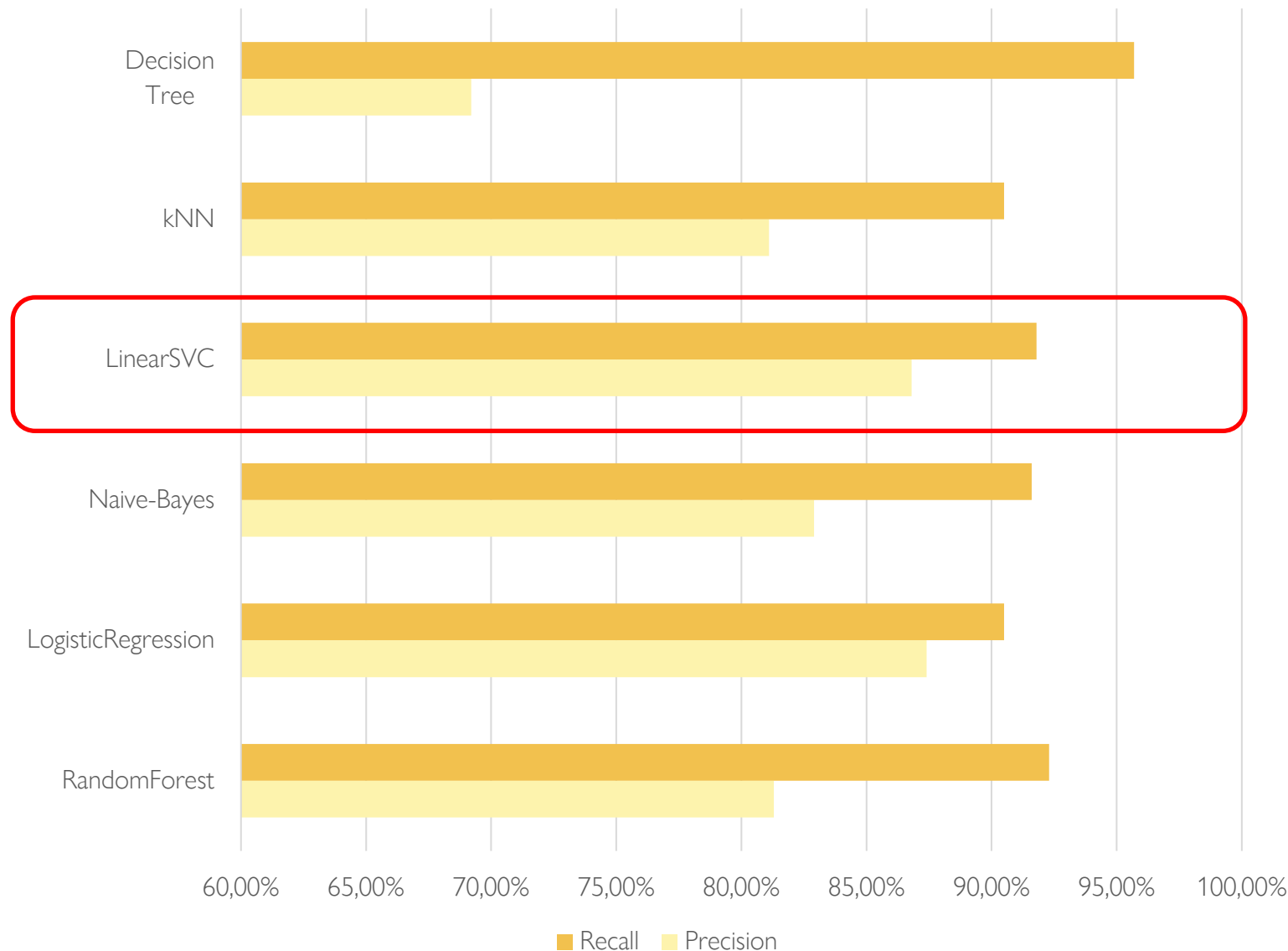
Precision: 71,5%

Recall: 90,5%

Risultati Classificatore Binario

Risultati

Risultati con Tfidf

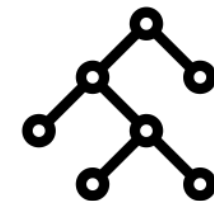
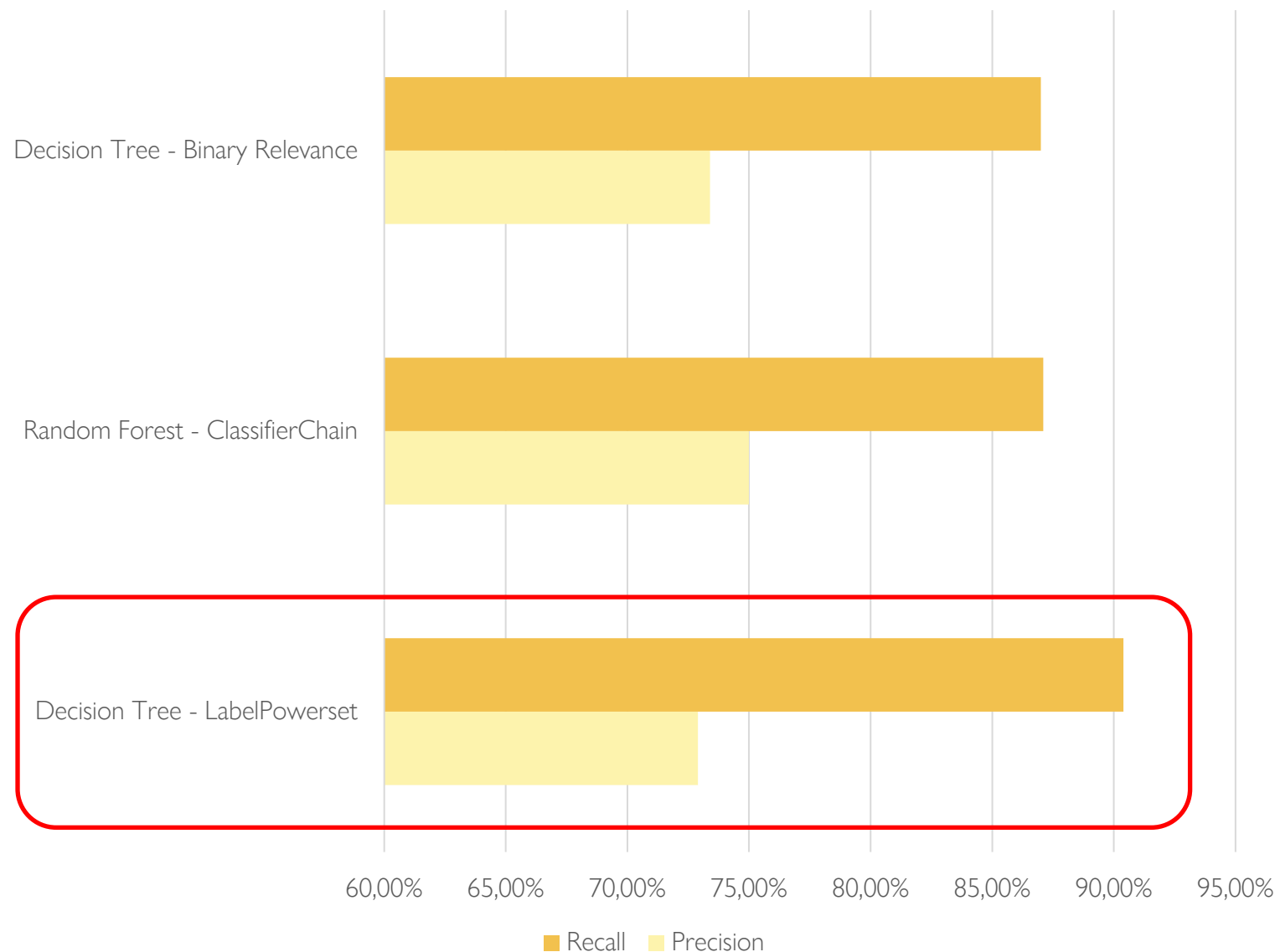


LinearSVC

Precision: 86,8%

Recall: 91,8%

Risultati con Tokenizer

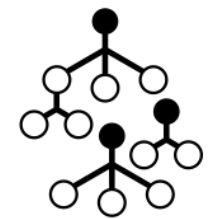
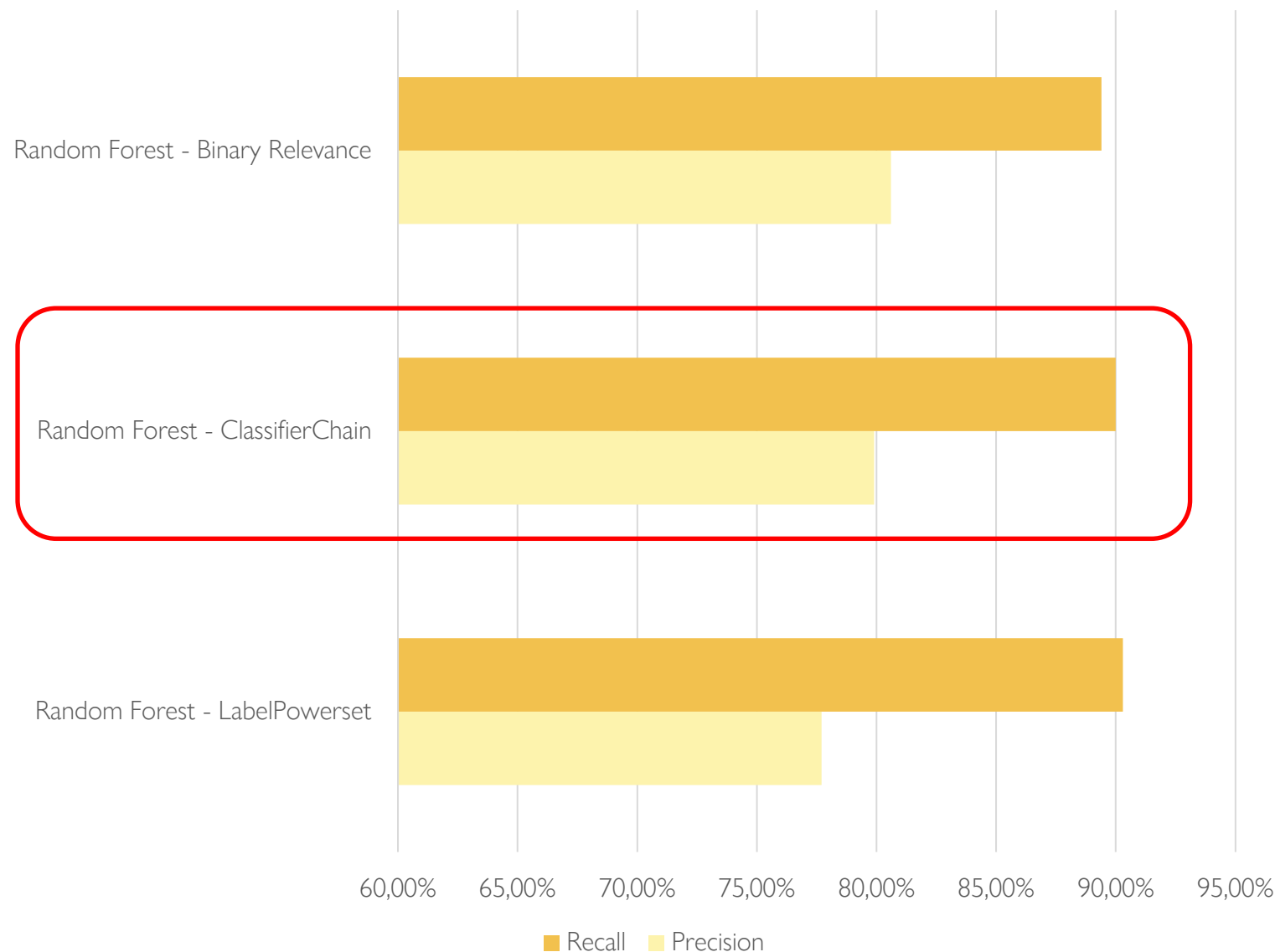


Decision Tree
LabelPowerset

Precision: 72,9%

Recall: 90,4%

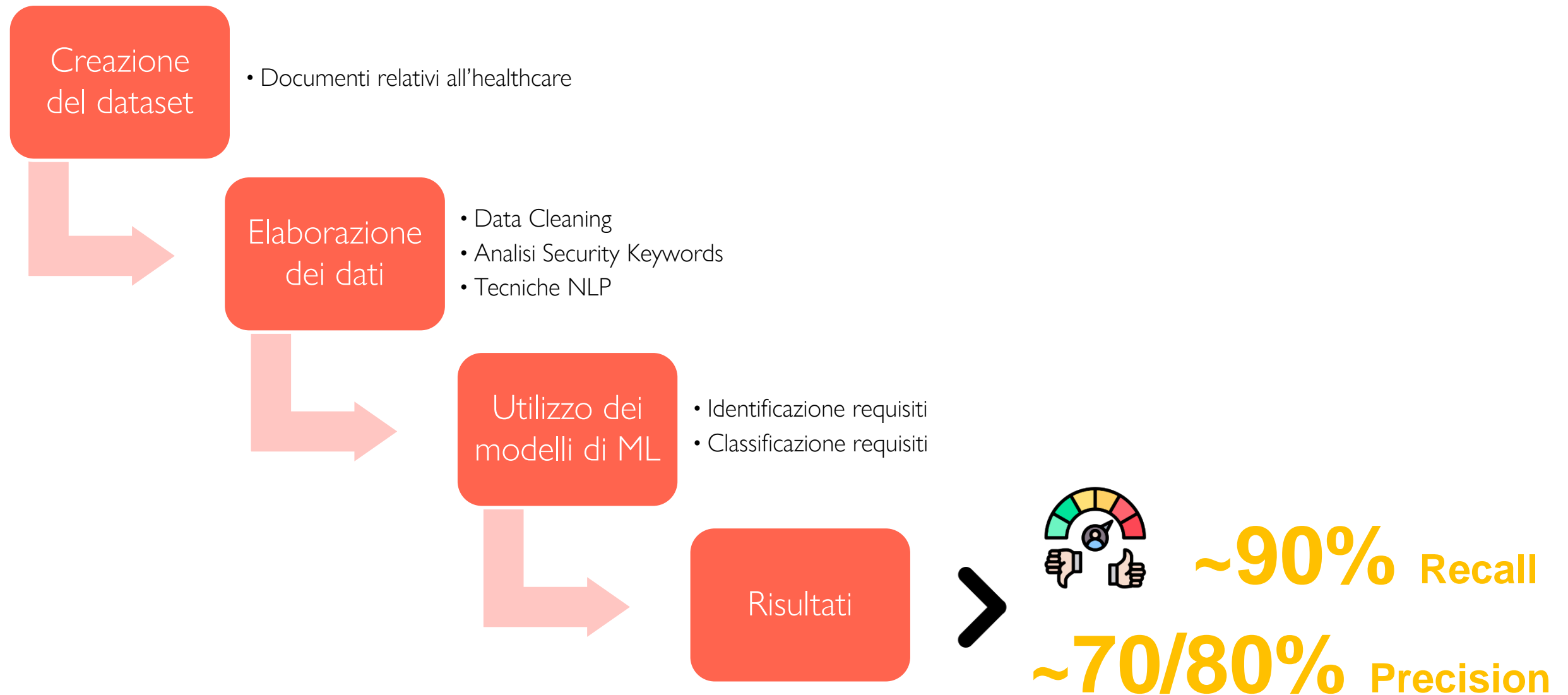
Risultati con Tfidf

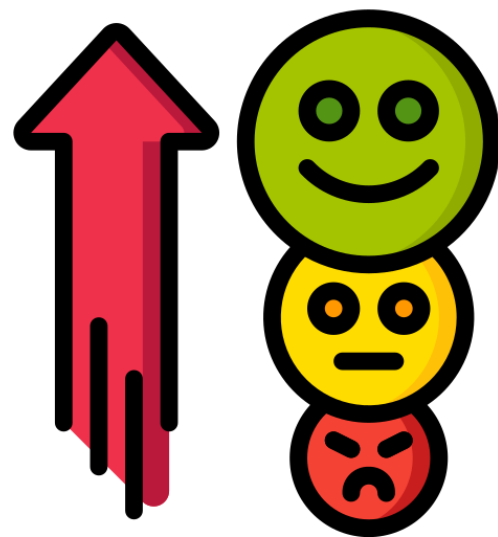


Random Forest
ClassifierChain

Precision: 79,9%

Recall: 90,0%





Miglioramenti futuri

- Dataset più generico e bilanciato
- Tecniche di Deep Learning
- Strumenti analoghi per altri NFR

Contesto

Introduzione e Background

sesa^{lab}
SOFTWARE ENGINEERING
RESEARCH



>

INGEGNERIA
DEI REQUISITI



REQUISITI
FUNZIONALI

REQUISITI NON
FUNZIONALI

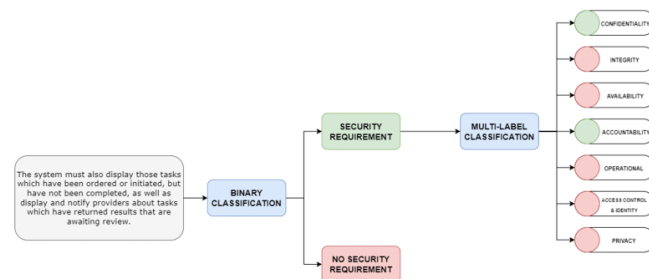
✉ d.mazza6@studenti.unisa.it
🌐 <https://github.com/xDaryamo>
🌐 <https://www.linkedin.com/in/dario-mazza-54b43a193/>

Estrazione di requisiti non funzionali di
sicurezza tramite tecniche di NLP
Dario Mazza

Architettura del progetto

Metodologia

sesa^{lab}
SOFTWARE ENGINEERING
RESEARCH



✉ d.mazza6@studenti.unisa.it
🌐 <https://github.com/xDaryamo>
🌐 <https://www.linkedin.com/in/dario-mazza-54b43a193/>

Estrazione di requisiti non funzionali di
sicurezza tramite tecniche di NLP
Dario Mazza

Risultati Classificatore Binario

Risultati

sesa^{lab}
SOFTWARE ENGINEERING
RESEARCH

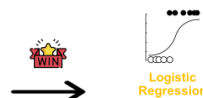
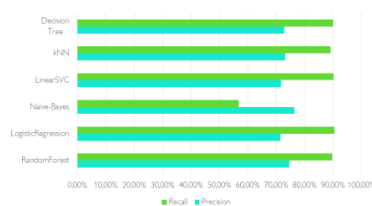
Precision $\frac{TP}{(TP + FP)}$

Recall $\frac{TP}{(TP + FN)}$

Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

F1-score: $2 \times \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

Risultati con Tokenizer



Precision: 71,5%

Recall: 90,5%

✉ d.mazza6@studenti.unisa.it
🌐 <https://github.com/xDaryamo>
🌐 <https://www.linkedin.com/in/dario-mazza-54b43a193/>

Estrazione di requisiti non funzionali di
sicurezza tramite tecniche di NLP
Dario Mazza

Conclusioni

sesa^{lab}
SOFTWARE ENGINEERING
RESEARCH



~90% Recall

~70/80% Precision



Miglioramenti futuri

- Dataset più generico e bilanciato
- Tecniche di Deep Learning
- Strumenti analoghi per altri NFR

✉ d.mazza6@studenti.unisa.it
🌐 <https://github.com/xDaryamo>
🌐 <https://www.linkedin.com/in/dario-mazza-54b43a193/>

Estrazione di requisiti non funzionali di
sicurezza tramite tecniche di NLP
Dario Mazza

Estrazione di requisiti non funzionali di sicurezza tramite tecniche di NLP

Grazie!



Questa tesi ha contribuito a
piantare un albero in Thailandia



Dario Mazza

d.mazza6@studenti.unisa.it
<https://github.com/xDaryamo>
[/in/dario-mazza-54b43a193/](https://www.linkedin.com/in/dario-mazza-54b43a193/)

