

Understanding Cybersecurity

A European and Italian Perspectivee

Project Work

Davide Piccirillo

Vincenzo Grimaldi

Miriam Varriale



Università Federico II di Napoli - Dipartimento di Scienze sociali
Corso di Laurea Magistrale in Sociologia Digitale e Analisi del web

Metodi Statistici per il data
mining



Analisi statistica delle net
communities

Cybersecurity: perché è importante ora più che mai

La Cybersecurity è un campo fondamentale della sicurezza informatica, volto a proteggere i sistemi da attacchi esterni, garantendo la confidenzialità, l'integrità e la disponibilità delle informazioni, insieme all'autenticità dei dati.

Le misure di sicurezza informatica riguardano aspetti giuridici, umani, tecnici e organizzativi, che analizzano le vulnerabilità, le minacce e i rischi associati.

Nel contesto aziendale, la protezione non si limita ai dati locali, ma si estende anche alla gestione dei flussi di dati tra dispositivi wireless, server cloud e la rete, con particolare attenzione ai dispositivi IoT e ai fornitori di terze parti.

L'obiettivo di questa analisi è comprendere il livello di sicurezza informatica delle imprese europee e italiane



Fondamenti della cybersecurity aziendale

- **Definizione:**
 - La sicurezza informatica aziendale protegge i dati e le risorse aziendali dalle minacce informatiche.
- **Protezione a livello locale e oltre:**
 - Utilizza metodi di sicurezza per proteggere i dati locali e ne estende la protezione al trasferimento di dati attraverso reti, dispositivi e utenti finali.
- **Minacce comuni:**
 - Attacchi DoS e DDoS (Denial-of-Service e Distributed Denial-of-Service)
 - Ingegneria sociale
 - Vulnerabilità del software
- **Sicurezza dei flussi di dati:**
 - Gestione sicura dei dati che vengono trasferiti tra dispositivi e reti all'interno dell'organizzazione.

Frame Teorico

Human Error Theory

La Teoria dell'Errore Umano esplora come il comportamento umano possa contribuire ai fallimenti dei sistemi, inclusi quelli legati alla sicurezza informatica. Secondo lo studio di Sasse, Brostoff e Weirich (2001), molti problemi di sicurezza derivano da comportamenti indesiderati degli utenti, spesso causati da:

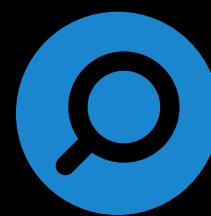
- **Mancanza di consapevolezza**: gli utenti potrebbero non comprendere l'importanza delle misure di sicurezza.
- **Procedure complesse o poco intuitive**: sistemi di sicurezza difficili da usare possono portare gli utenti a evitarli o aggirarli.
- **Formazione insufficiente**: la mancanza di addestramento adeguato può aumentare la probabilità di errori.

Risk Management Theory

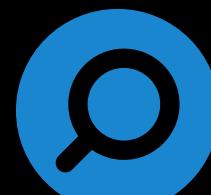
La Gestione del Rischio è un processo sistematico attraverso il quale le organizzazioni identificano, analizzano e affrontano i rischi che potrebbero influenzare il raggiungimento dei loro obiettivi. Secondo la norma ISO 31000:2018, questo processo comprende:

- **Identificazione dei rischi**: riconoscere cosa potrebbe accadere e come potrebbe influenzare gli obiettivi dell'organizzazione.
- **Valutazione dei rischi**: comprendere la natura dei rischi e determinarne il livello.
- **Trattamento dei rischi**: selezionare e implementare opzioni per affrontare i rischi.
- **Monitoraggio e revisione**: osservare e valutare periodicamente il processo di gestione del rischio per garantire la sua efficacia.

Interrogativi di ricerca

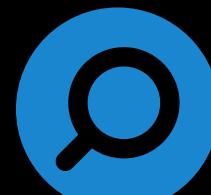


Qual è il livello di adozione delle misure di sicurezza ICT tra le imprese nei diversi paesi europei?



Qual è il livello di adozione delle misure di sicurezza ICT delle imprese italiane?

È possibile individuare gruppi di imprese italiane con comportamenti simili nella gestione della sicurezza ICT?



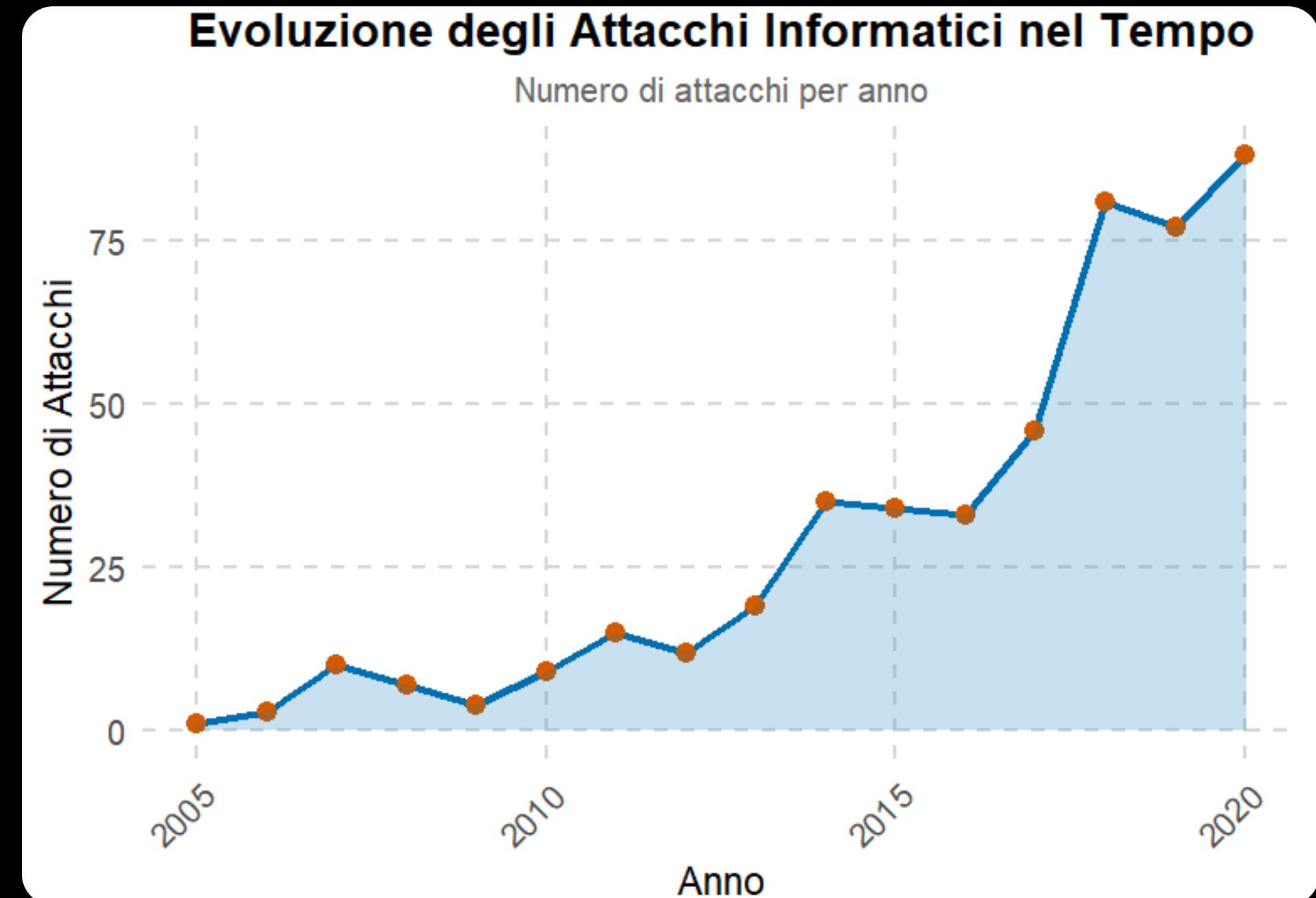
Come le imprese e le infrastrutture critiche europee affrontano le minacce alla cybersecurity provenienti da attori esterni e geopolitici?

Un primo sguardo

Dal 2005 al 2020, il numero di attacchi informatici è aumentato in modo significativo. Nei primi anni gli attacchi erano relativamente pochi, ma il trend mostra un'accelerazione costante.

La crescita degli attacchi è parallela all'espansione delle tecnologie digitali. Con la diffusione di Internet, cloud computing e smart devices, le superfici di attacco sono aumentate.

Il trend evidenzia come sia sempre più critico proteggere dati e infrastrutture. Le aziende devono adottare misure di sicurezza avanzate per ridurre il rischio di attacchi.



Analisi condotta sul dataset “Cyber incidents from 2005 to 2020”

Fonte - Kaggle/CFR

Prima analisi

Livello europeo

La cybersecurity nelle imprese europee sta assumendo un'importanza sempre maggiore con l'evoluzione delle tecnologie e l'aumento delle minacce informatiche e dei device sempre più connessi



I dati raccolti da vari paesi dell'Unione Europea offrono uno spunto per capire come le misure di sicurezza ICT vengano adottate a livello aziendale e come le imprese affrontino i rischi informatici.

Metadati principali

- **Fonte:** Eurostat - ICT Security In Enterprise
 - **Anno di riferimento:** 2021/2022
 - **Unità di analisi:** Imprese europee (dati aggregati per paese)
 - **Formato:** Excel
 - **Copertura geografica:** Paesi UE + 5 extra-UE
 - **Tipo di dati:** Percentuali

	Use at least one ICT security measure	Make persons employed aware of their ob...
Country	ICT_Measure	Value
EU		92
Belgium		96
Bulgaria		82
Czechia		92
Denmark		98
Germany		96
Estonia		86
Ireland		91
Greece		75
Spain		88
France		93
Croatia		85
Italy		92
Cyprus		91
Latvia		83
Lithuania		88
Luxembou		86
Hungary		79
Malta		93
Netherland		96
Austria		92
Poland		93
Portugal		90
Romania		86
Slovenia		87
Slovakia		86
Finland		98
Sweden		91



Pre-trattamento

- Trasformato i quesiti in variabili
 - Rimossi i valori mancanti
 - Filtrato le variabili per la generazione dei grafici di interesse

A large, solid blue arrow pointing to the right, centered on the page. It serves as a visual cue to guide the user's eye towards the search results listed on the right side of the interface.



Ogni osservazione del dataset rappresenta un paese europeo per un totale di 32 osservazioni (compresi paesi extra UE come Montenegro, Serbia, Turchia) e 6 variabili numeriche associate

Variabili nel dataset:

1. **ICT_Measures**
2. **Employee_Awareness**
3. **ICT_Documents**
4. **ICT_Insurance**
5. **ICT_Review**
6. **ICT_Incidents_2021**

Country	ICT_Measures	Employee_Awareness	ICT_Documents	ICT_Insurance	ICT_Review	ICT_Incidents_2021
Belgium	96	54	34	33	24	23
Bulgaria	82	48	22	4	17	11
Czechia	92	75	26	12	18	29
Denmark	98	70	55	71	37	26
Germany	96	68	36	32	24	26
Estonia	86	61	37	9	21	26
Ireland	91	75	51	42	35	15
Greece	75	32	18	14	7	18
Spain	88	51	29	22	19	17
France	93	47	21	40	12	26
Croatia	85	38	49	7	29	19
Italy	92	62	48	16	30	16
Cyprus	91	52	31	17	21	14
Latvia	83	50	49	9	40	19
Lithuania	88	63	34	5	17	16
Luxembourg	86	52	26	31	19	16
Hungary	79	43	33	5	21	13
Malta	93	60	38	35	26	26
Netherlands	96	46	44	30	34	30
Austria	92	61	32	27	19	15
Poland	93	51	44	14	21	30
Portugal	90	63	54	11	34	11
Romania	86	62	45	7	40	19
Slovenia	87	51	40	8	27	15
Slovakia	86	62	25	10	17	12
Finland	98	67	57	33	43	44
Sweden	91	65	66	46	37	21
Norway	91	58	30	38	21	17
Montenegro	62	51	15	17	10	23
Serbia	96	67	55	12	36	15
Türkiye	76	69	35	7	23	29
Bosnia and Herzegovina	86	46	26	13	16	21

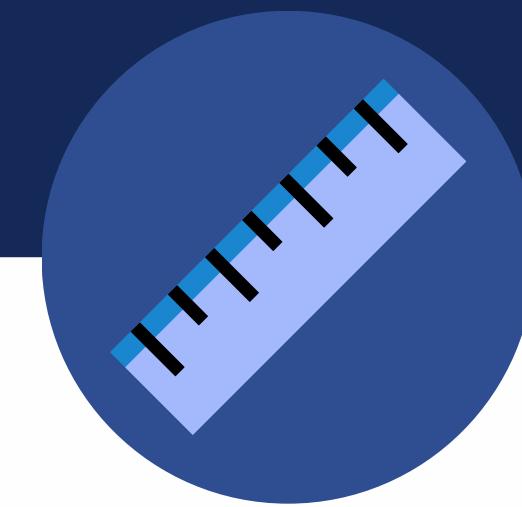
Statistiche Descrittive

Variabili	Minimo	Massimo	Campo di variazione	Media	Deviazione standard	Varianza
ICT_Incidents_2021	11	44	33	20	7	52
ICT_Measures	62	98	36	88	7	57
ICT_Insurance	4	71	67	21	15	241
ICT_Documents	15	66	51	38	13	160
ICT_Review	7	43	36	25	9	89
Employee_Awareness	32	75	43	57	10	109



ICT Review

- Solo il 24.8% delle imprese europee aggiorna i documenti ICT almeno annualmente.
- Il range ampio (36%) mostra che alcuni paesi sono molto diligenti, altri quasi per nulla.
- La disomogeneità tra paesi (deviazione 9.37%) suggerisce una carenza di regolamentazione o prassi comuni.



ICT Measures

- L'adozione di misure ICT è molto diffusa: quasi tutte le imprese in ogni paese ne hanno almeno una.
- La bassa deviazione mostra un livello di sicurezza minimo piuttosto uniforme in tutta Europa.
- Questa è la variabile più consolidata e omogenea.



ICT Insurance

- Grande disparità tra paesi: ci sono alcuni con pochissime assicurazioni, altri con molte.
- La media bassa (21%) mostra che la copertura assicurativa è ancora marginale.
- L'alta varianza rivela che la gestione del rischio è molto frammentata.



ICT Incidents 2021

- In media, 1 azienda su 5 ha subito un incidente grave legato all'ICT.
- Paesi con valori molto alti (fino al 44%) potrebbero avere debolezze sistemiche o una maggiore digitalizzazione non protetta.
- La variabilità (dev. std 7.21%) indica che alcuni paesi sono più vulnerabili di altri.



ICT Documents

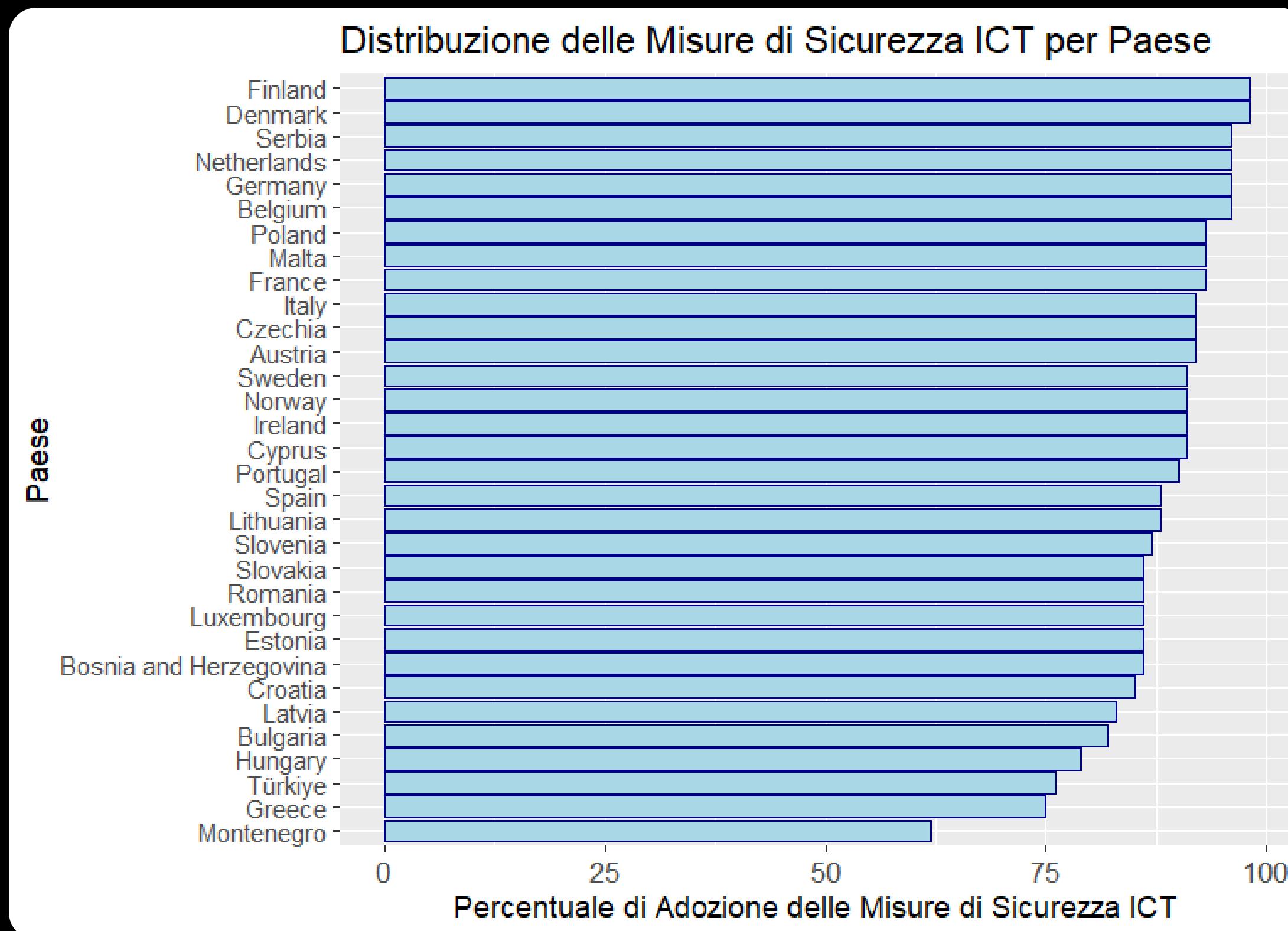
- La documentazione formale è presente in meno della metà delle imprese, mediamente.
- Il campo molto ampio (51%) evidenzia una forte diversità di pratiche, potenzialmente dovuta a regolamenti nazionali differenti.



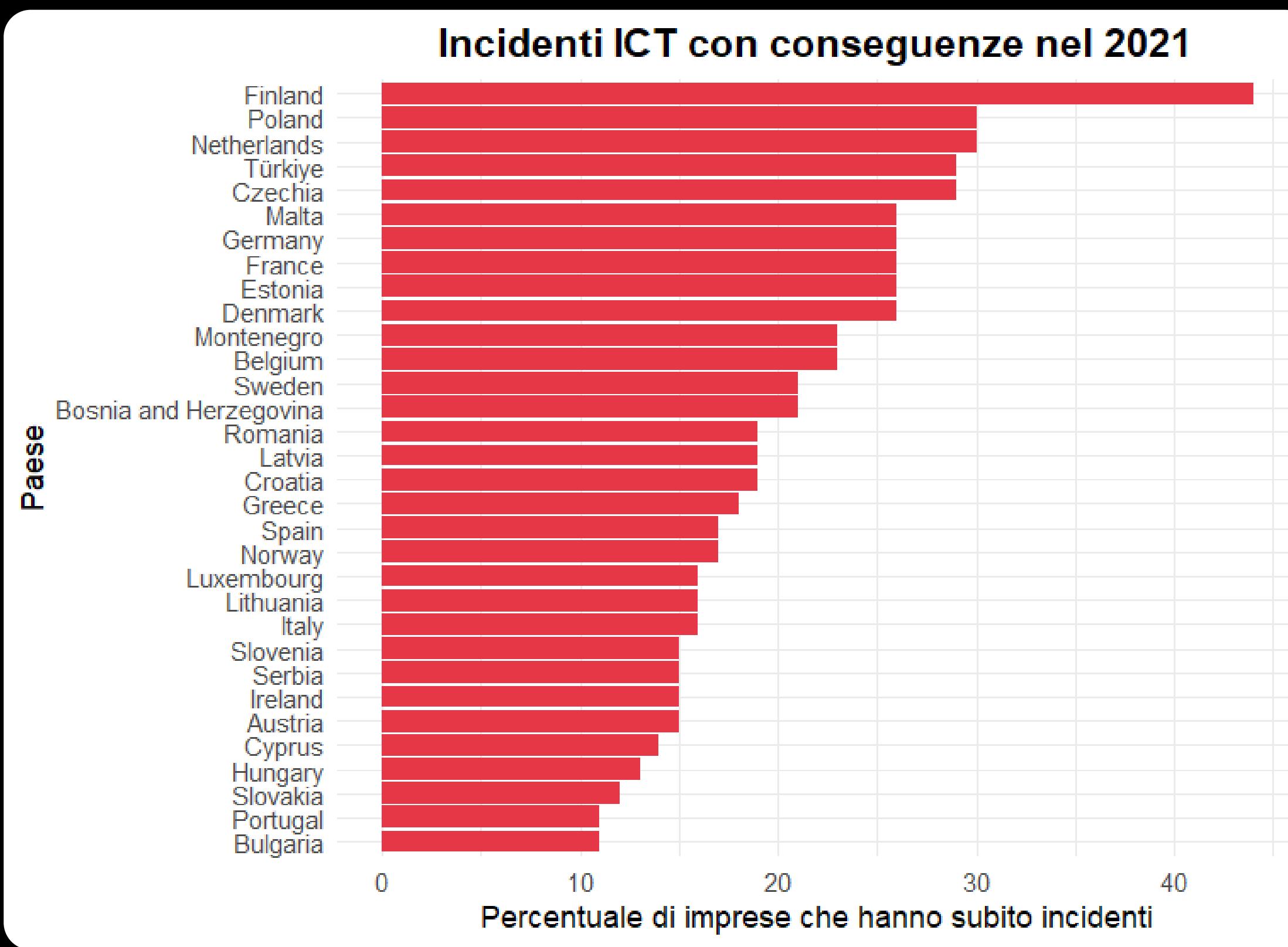
Employee Awareness

- In media più della metà delle imprese sensibilizza i dipendenti.
- Tuttavia, in alcuni paesi quasi 7 aziende su 10 lo fanno, in altri meno di un terzo.
- La consapevolezza dei dipendenti è essenziale nella prevenzione degli attacchi (phishing, errori umani ecc.)

Analisi delle misure di sicurezza ICT: Distribuzione tra paesi



Analisi degli incidenti ICT: Distribuzione tra paesi



Seconda analisi

A dark blue background featuring a central, glowing blue wave pattern composed of numerous small dots, resembling a digital or quantum wave.

Livello Italiano

Da una prima analisi europea emerge che l'Italia si colloca in una posizione intermedia rispetto al resto dei paesi, con una buona adozione delle misure di sicurezza ICT e un numero di incidenti informatici relativamente basso.

Tuttavia, nonostante un'adeguata formazione del personale, il livello di copertura assicurativa contro gli incidenti ICT resta molto basso. Anche la presenza di documentazione formale sulla sicurezza risulta carente, indicando la necessità di attuare un approccio più strutturato e robusto

Per approfondire il contesto italiano, è stata svolta un'ulteriore analisi su un secondo dataset, focalizzato sui settori industriali e le relative pratiche di cybersecurity.



Employee_Awareness	ICT_Insurance	ICT_Documents
62	16	48

Metadati principali

- **Fonte:** ISTAT - Imprese e ICT
- **Anno di riferimento:** 2021/2022
- **Unità di analisi:** Settori d'impresa Italiane
- **Formato:** Excel
- **Copertura geografica:** Italia
- **Tipo di dati:** Percentuali

Pre-trattamento

- Rinominato le variabili con etichette più leggibili.
- Trasformato le variabili in formato numerico.
- Eliminato i valori mancanti.
- Eseguita l'analisi descrittiva degli indicatori quantitativi.

Attività	Indicatore1	Indicatore2	Indicatore3
	Autenticazione con password complessa	Backup dei dati	Identificazione e autenticazione dei transatti
Attività manifatturiera	84	85	
Industrie alimentari, delle bevande e del tabacco	83	78	
Industrie tessili, dell'abbigliamento, articoli in pelle e simili	82	75	
Industria dei prodotti in legno e carta, stampa			
Fabbricazione di coke e di prodotti derivanti dalla raffinazione del petrolio, di prodotti chimici, di prodotti farmaceutici, di articoli in gomma e materie plastiche e di prodotti della lavorazione di minerali non metalliferi			
Metallurgia e fabbricazione di prodotti in metallo esclusi macchinari e attrezzature			
Fabbricazione di computer e prodotti di elettronica e ottica, apparecchi			
Costruzioni	89.96	77.40	
Alloggio	83.37	81.80	
Altre industrie manifatturiere, riparazione e installazione di macchinari e attrezzature	92.89	94.36	
Attività dei servizi delle agenzie di viaggio, dei tour operator...	93.67	85.06	
Attività di produzione cinematografica, di video e di programmi televisivi	90.36	94.74	
Attività editoriali	78.48	80.51	
Attività immobiliari	84.08	84.94	
Attività manifatturiere	91.46	91.38	
Attività professionali, scientifiche e tecniche	76.26	68.61	
Commercio al dettaglio (escluso quello di autoveicoli e di macchinari)	83.36	79.73	
Commercio all'ingrosso e al dettaglio, riparazione di autoveicoli e macchinari	83.34	79.36	
Fabbricazione di apparecchiature elettriche ed apparecchiature di telecomunicazioni	85.94	89.60	
Fabbricazione di coke e di prodotti derivanti dalla raffinazione del petrolio, di prodotti chimici, di prodotti farmaceutici, di articoli in gomma e materie plastiche e di prodotti della lavorazione di minerali non metalliferi	82.33	85.18	
Fabbricazione di computer e prodotti di elettronica e ottica, apparecchi	93.77	96.26	
Fabbricazione di mezzi di trasporto	94.83	95.55	
Fornitura di energia elettrica, gas, vapore e aria condizionata	88.03	85.15	
Industria dei prodotti in legno e carta, stampa	78.88	85.27	
Industrie alimentari, delle bevande e del tabacco	83.27	77.63	
Industrie tessili, dell'abbigliamento, articoli in pelle e simili	82.19	74.70	
Informatica ed altri servizi d'informazione	95.31	92.30	
Metallurgia e fabbricazione di prodotti in metallo esclusi macchinari e attrezzature	85.23	91.35	

Nel dettaglio, il dataset è composto da:

- **26 osservazioni**, ciascuna corrispondente a un settore d'impresa;
- **13 variabili**, che rappresentano diversi indicatori del livello di preparazione delle imprese in ambito cybersecurity.

Gli indicatori considerano un ampio spettro di misure di sicurezza, che spaziano dalle soluzioni più tradizionali fino a quelle più avanzate, come l'impiego di tecnologie biometriche.

È inoltre presente la componente umana: **due variabili** del dataset si riferiscono infatti alla formazione dei dipendenti in materia di sicurezza informatica, riconoscendo l'importanza del fattore umano nella protezione dei sistemi aziendali.

	Attività	Pass.compl	Backup.d	Biometric	Crittogr	Acc.rete	VPN	Reg.file	Risk.value	E
1	Alloggio	89.96	77.40	4.85	28.59	64.93	40.48	52.13	31.65	
2	Altre industrie manifatturiere, riparazione e installazione di ...	83.37	81.80	6.66	17.67	59.16	37.37	40.09	37.34	
3	Attività dei servizi delle agenzie di viaggio, dei tour operator...	92.89	94.36	4.67	33.71	88.06	68.81	67.21	56.89	
4	Attività di produzione cinematografica, di video e di programma...	93.67	85.06	13.41	27.29	73.51	47.93	45.14	29.05	
5	Attività editoriali	90.36	94.74	12.78	33.04	84.38	68.13	66.30	65.02	
6	Attività immobiliari	78.48	80.51	4.42	26.34	67.41	41.92	47.81	46.83	
7	Attività manifatturiere	84.08	84.94	7.24	20.02	66.63	45.25	47.43	37.35	
8	Attività professionali, scientifiche e tecniche	91.46	91.38	7.29	30.33	84.22	64.31	63.32	52.91	
9	Commercio al dettaglio (escluso quello di autoveicoli e di m...	76.26	68.61	6.13	14.94	52.01	28.72	31.47	27.45	
10	Commercio all'ingrosso e al dettaglio, riparazione di autovei...	83.36	79.73	6.92	20.55	62.86	41.38	43.90	35.76	
11	Costruzioni	83.34	79.36	10.09	17.11	55.59	32.05	36.82	25.56	
12	Fabbricazione di apparecchiature elettriche ed apparecchiatur...	85.94	89.60	9.79	22.84	71.37	61.42	54.62	48.08	
13	Fabbricazione di coke e di prodotti derivanti dalla raffinazio...	82.33	85.18	6.24	23.99	74.46	56.29	53.15	45.76	
14	Fabbricazione di computer e prodotti di elettronica e ottica, ...	93.77	96.26	10.00	36.53	84.99	75.95	64.38	47.66	
15	Fabbricazione di mezzi di trasporto	94.83	95.55	9.74	21.75	82.83	35.66	53.95	48.47	
16	Fornitura di energia elettrica, gas, vapore e aria condizionat...	88.03	85.15	8.39	29.52	70.97	53.24	52.16	43.75	
17	Industria dei prodotti in legno e carta, stampa	78.88	85.27	3.39	22.38	70.84	43.93	52.68	34.91	
18	Industrie alimentari, delle bevande e del tabacco	83.27	77.63	6.06	20.79	56.38	35.55	45.68	28.09	
19	Industrie tessili, dell'abbigliamento, articoli in pelle e simili	82.19	74.70	8.21	14.73	53.29	32.71	38.25	25.44	
20	Informatica ed altri servizi d'informazione	95.31	92.30	13.92	50.38	86.61	80.17	76.76	67.40	
21	Metallurgia e fabbricazione di prodotti in metallo esclusi m...	85.23	91.35	6.97	17.93	72.30	44.55	46.35	36.81	

Statistiche Descrittive

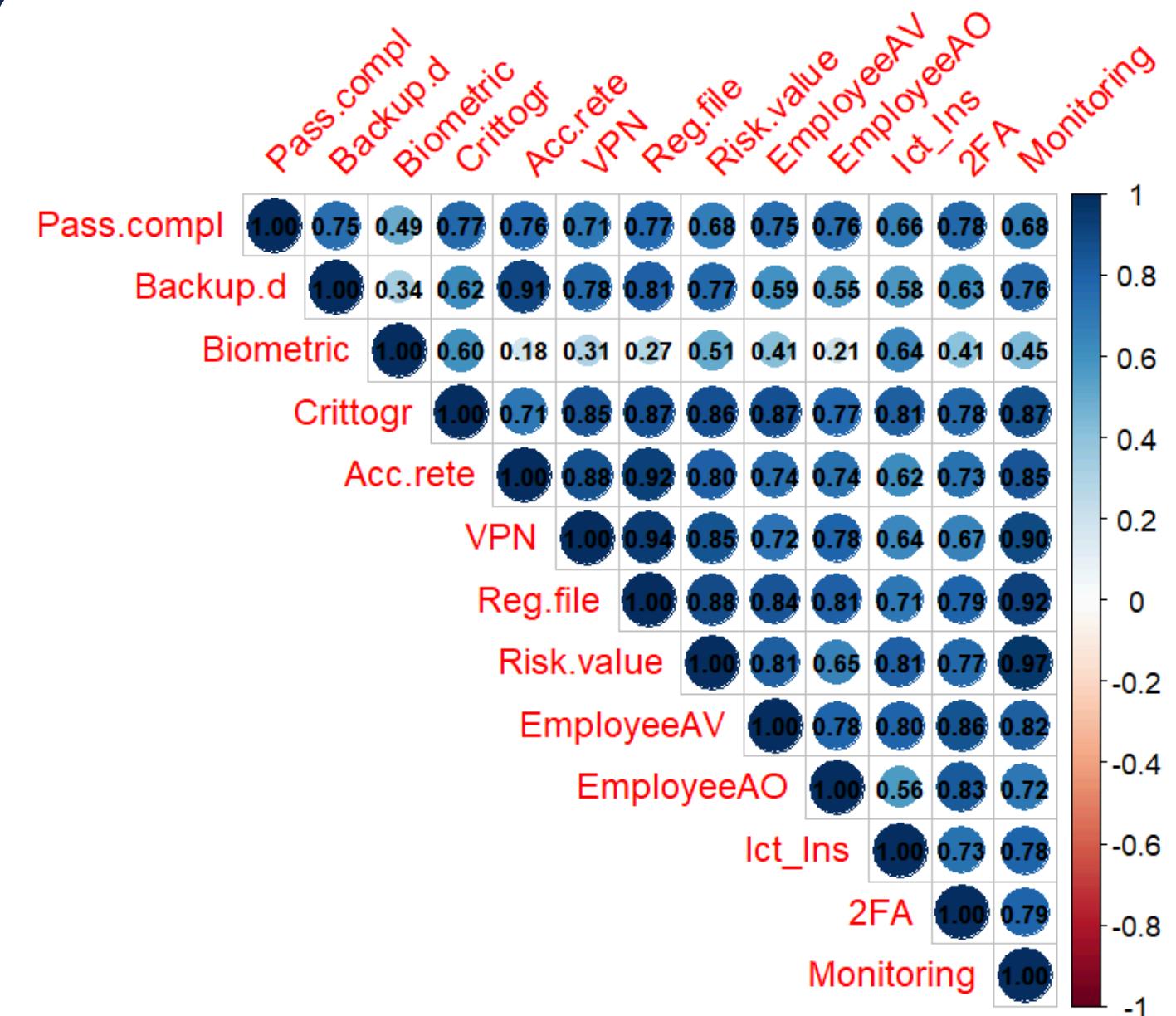
Variabili	Minimo	Massimo	Campo di variazione	Media	Deviazione standard	Varianza
Password complesse	76,26	96,58	20,32	87,28	5,85	34,25
Backup di dati	57,10	96,26	39,16	84,28	9,27	85,84
Misure Biometriche	3,39	33,55	30,16	9,86	6,24	38,97
Tecniche di crittografia	14,73	53,66	38,93	27,37	10,65	113,40
Controllo accesso rete	28,65	90,38	61,73	68,75	14,79	218,85
VPN (Virtual Private Network)	21,66	84,51	62,85	49,49	17,50	306,41

Variabili	Minimo	Massimo	Campo di variazione	Media	Deviazione standard	Varianza
File di registro (logfile)	25,12	77,64	52,52	51,31	13,50	182,38
Valutazione del rischio	13,79	74,37	60,58	42,71	14,66	214,88
Dipendenti formati volontariamente	27,09	80,77	53,58	51,09	12,25	150,03
Dipendenti formati obbligatoriamente	6,78	57,18	50,40	26,20	11,14	124,01
Assicurazione ICT	7,36	50,35	42,99	21,82	11,15	124,29
Autenticazione a 2 fattori (2FA)	22,17	48,57	26,40	29,92	6,66	44,34
Uso di strumenti di monitoring	16,29	76,27	59,98	42,18	14,60	213,25

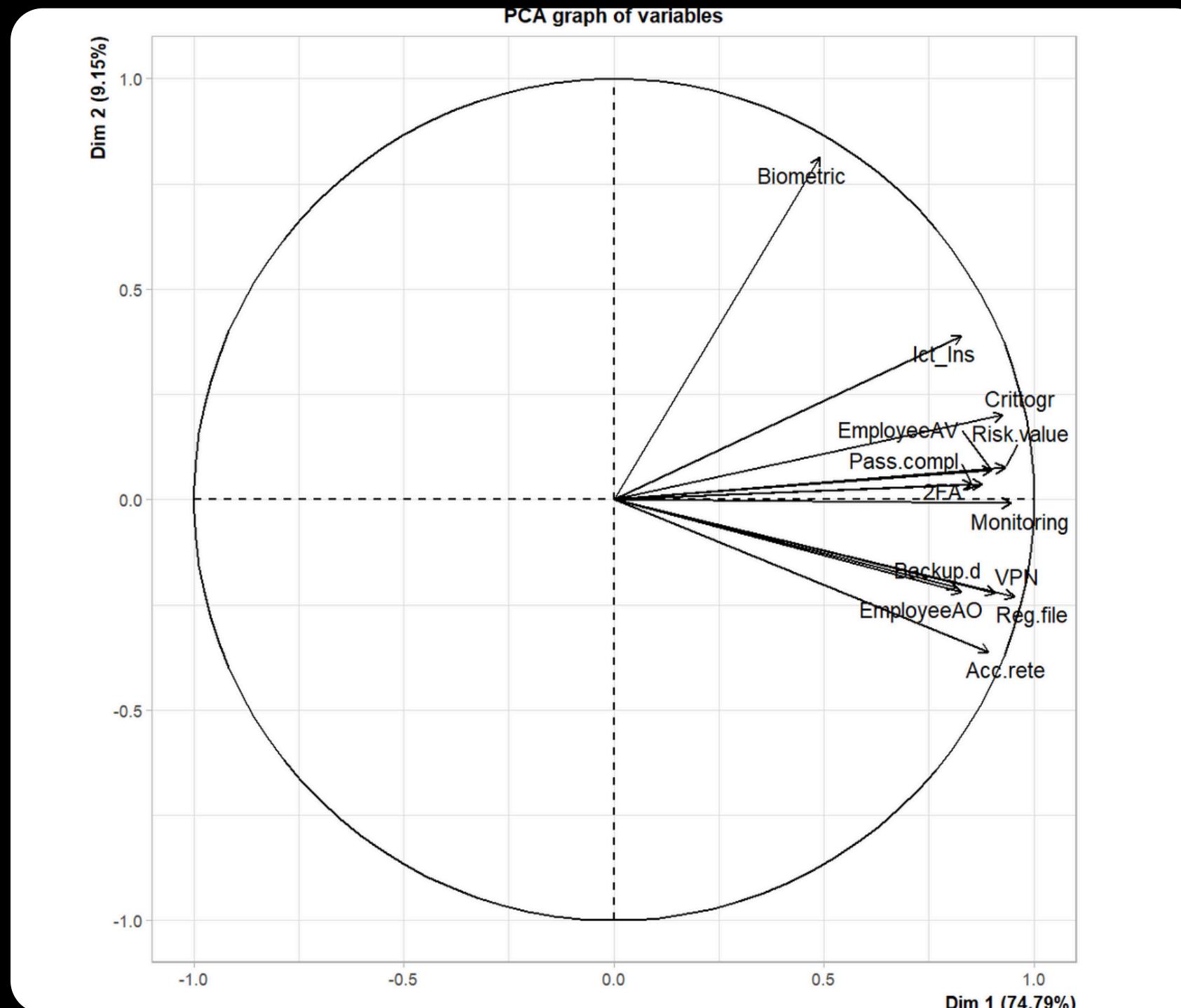
Analisi delle Correlazioni tra le Misure di Sicurezza ICT

Per verificare la presenza di relazioni tra gli **indicatori di sicurezza ICT**, è stata calcolata la matrice di correlazione tra le variabili numeriche.

- Forte correlazione tra la maggior parte delle variabili (es. Reg.file, ICT.test, VPN, Risk.value, Crittogr, Acc.rete, 2FA, **tutte > 0.85**): ciò indica la presenza di un comportamento coerente nei settori che adottano strumenti di sicurezza – Chi implementa una misura, tende a implementarle quasi tutte.
 - La variabile Biometric mostra correlazioni molto basse con tutte le altre (**massimo ≈ 0.34**): suggerisce che rappresenta una dimensione distinta, legata a tecnologie avanzate adottate da **pochi settori specifici**.



ACP - Analisi Componenti Principali



La variabilità spiegata dai fattori:

- Primo fattore (Dim. 1): 74,8% della variabilità totale;
- Secondo fattore (Dim. 2): 9,2% della variabilità totale;

I due fattori insieme riescono a spiegare l'84% della varianza totale.

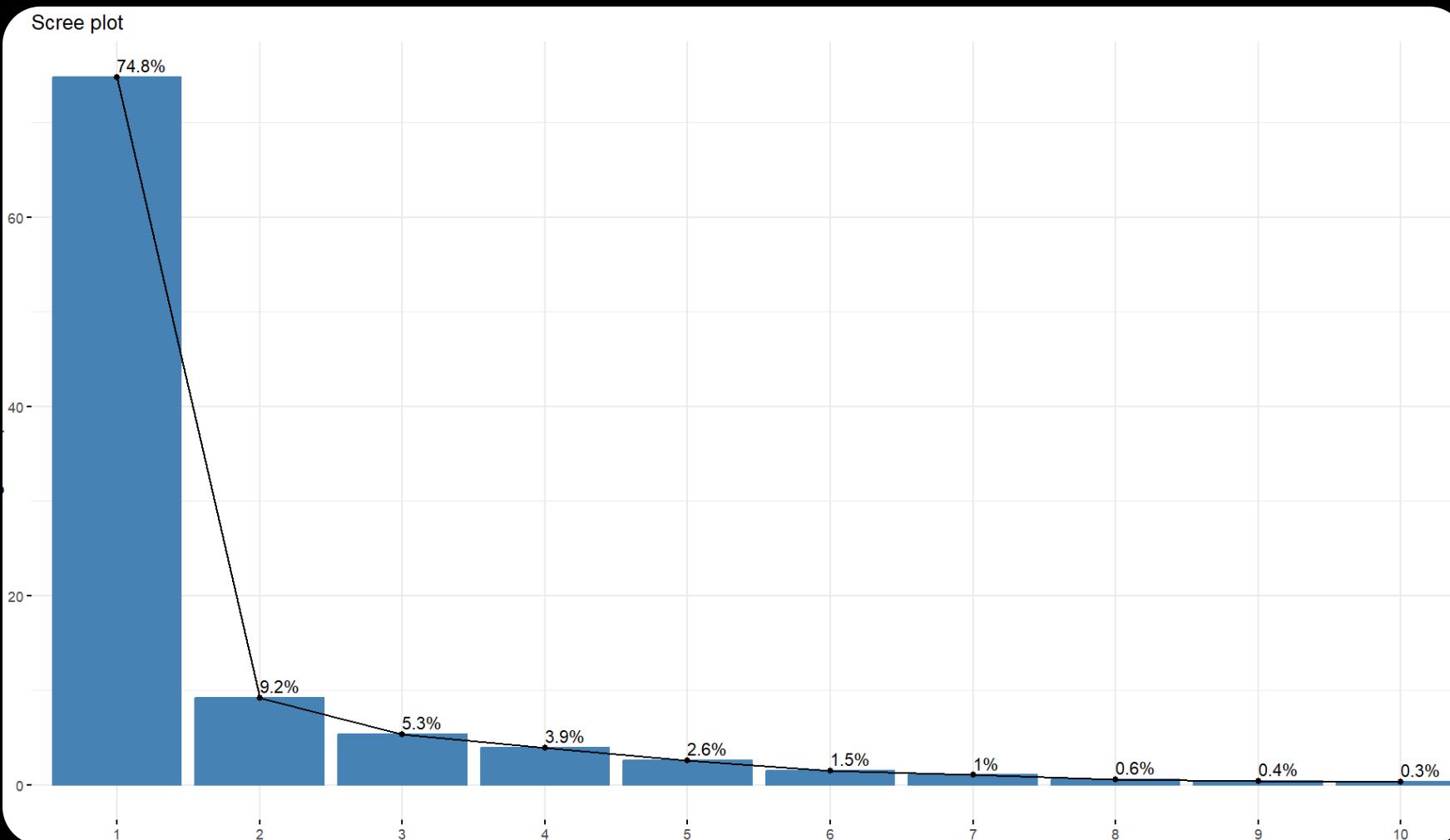
Le variabili molto correlate alla **Dim.1** sono:

- **Reg.File (0.953);**
- **Monitoring (0.945);**
- **Risk.Value (0.931);**
- **Crittogr. (0.926);**
- **VPN (0.907);**
- **EmployeeAV (0.899)**

La variabile molto correlata alla **Dim. 2** è:

- **Biometric (0.813)**

Criteri di scelta delle dimensioni: Autovalori e Screeplot



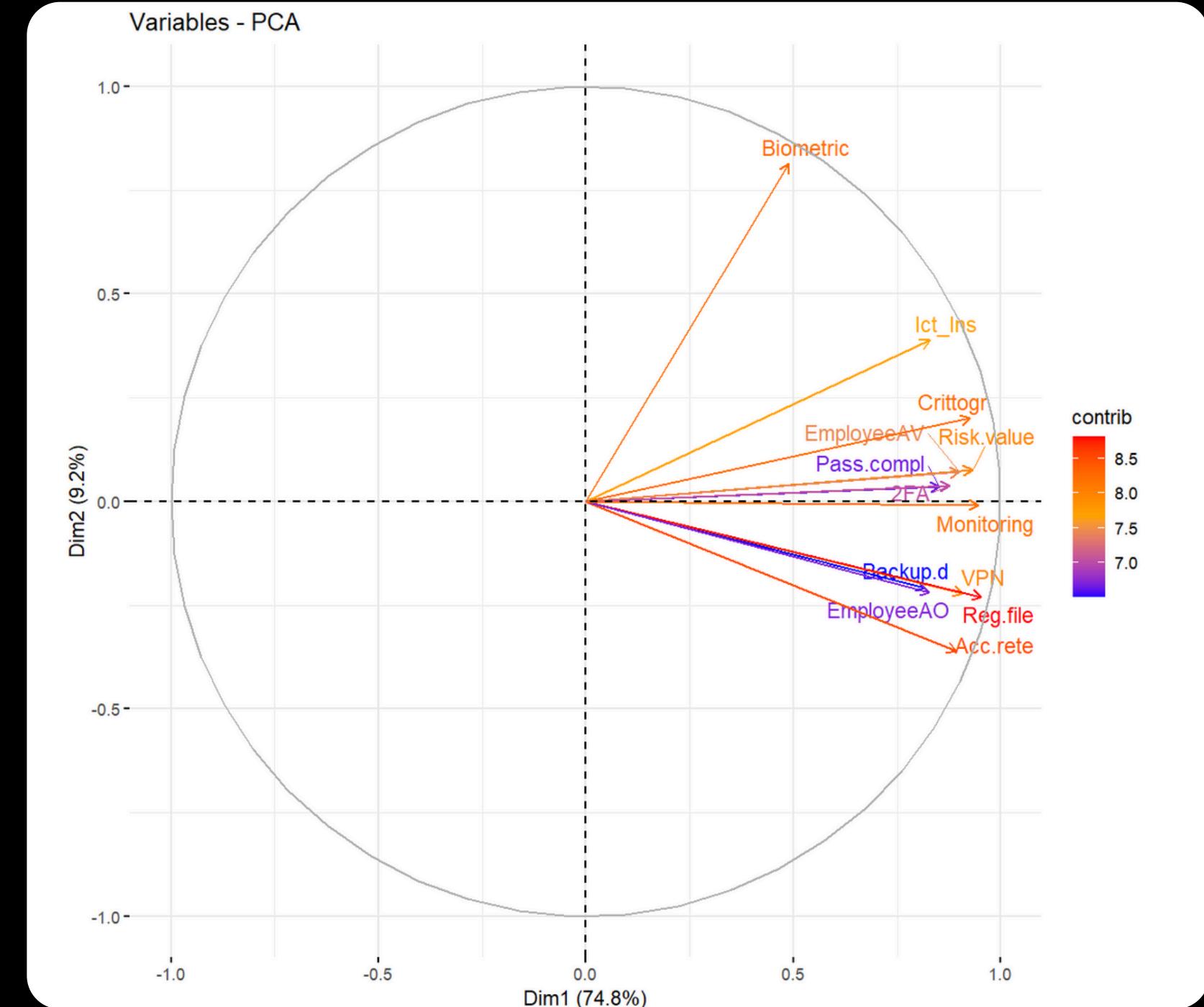
	eigenvalue	variance.percent	cumulative.variance.percent
Dim. 1	9.72	74.79	74.79
Dim. 2	1.19	9.15	83.95
Dim. 3	0.69	5.34	89.28
Dim. 4	0.51	3.93	93.21
Dim. 5	0.34	2.59	95.80
Dim. 6	0.20	1.50	97.30
Dim. 7	0.13	1.01	98.32
Dim. 8	0.07	0.56	98.88
Dim. 9	0.05	0.40	99.28

L'analisi degli autovalori e lo screeplot confermano che l'informazione contenuta nel dataset può essere sintetizzata efficacemente su due componenti principali.

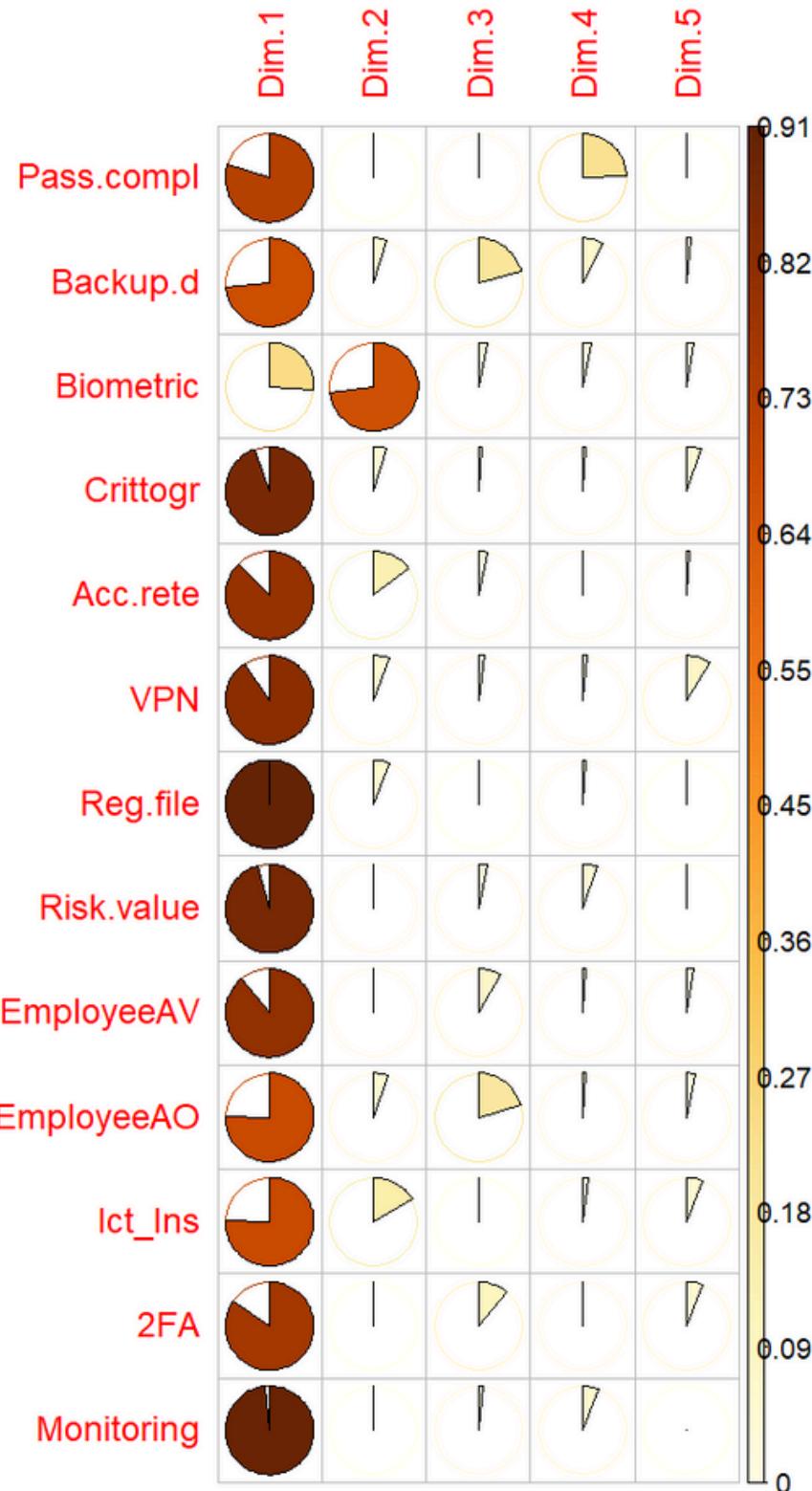
- La prima componente (**Dim.1**) ha un autovalore pari a 9,72 e spiega da sola il 74,8% della varianza totale, rappresentando un asse dominante che cattura la maggior parte delle differenze tra i settori.
- La seconda componente (**Dim.2**) ha un autovalore di 1,19 e aggiunge un ulteriore 9,2% di varianza spiegata, contribuendo a evidenziare ulteriori distinzioni più sottili.

Contributo delle variabili alle componenti principali: Cerchio delle Correlazioni

- **Variabili più rilevanti (rosse, con freccia lunga e ben orientata)**
 - Acc.rete, Reg.file, EmployeeAO, VPN, Backup.d, Risk.value, Crittogr, Ict_Ins, Biometric.
 - Hanno forti contributi su Dim.1
 - Mostrano che Dim.1 rappresenta una dimensione condivisa e strutturata di adozione delle misure ICT
- **Biometric**
 - Posizionata quasi verticale, molto lontana dalle altre
 - molto correlata alla Dim.2
 - È la variabile che “spiega” quasi da sola la seconda componente
- **Variabili con contributo minore**
 - Monitoring, EmployeeAV, Pass.compl
 - Frecce leggermente più corte o colori meno intensi (giallo/viola). Contribuiscono meno alla definizione delle prime due dimensioni mostrate



Qualità della rappresentazione: Cos²

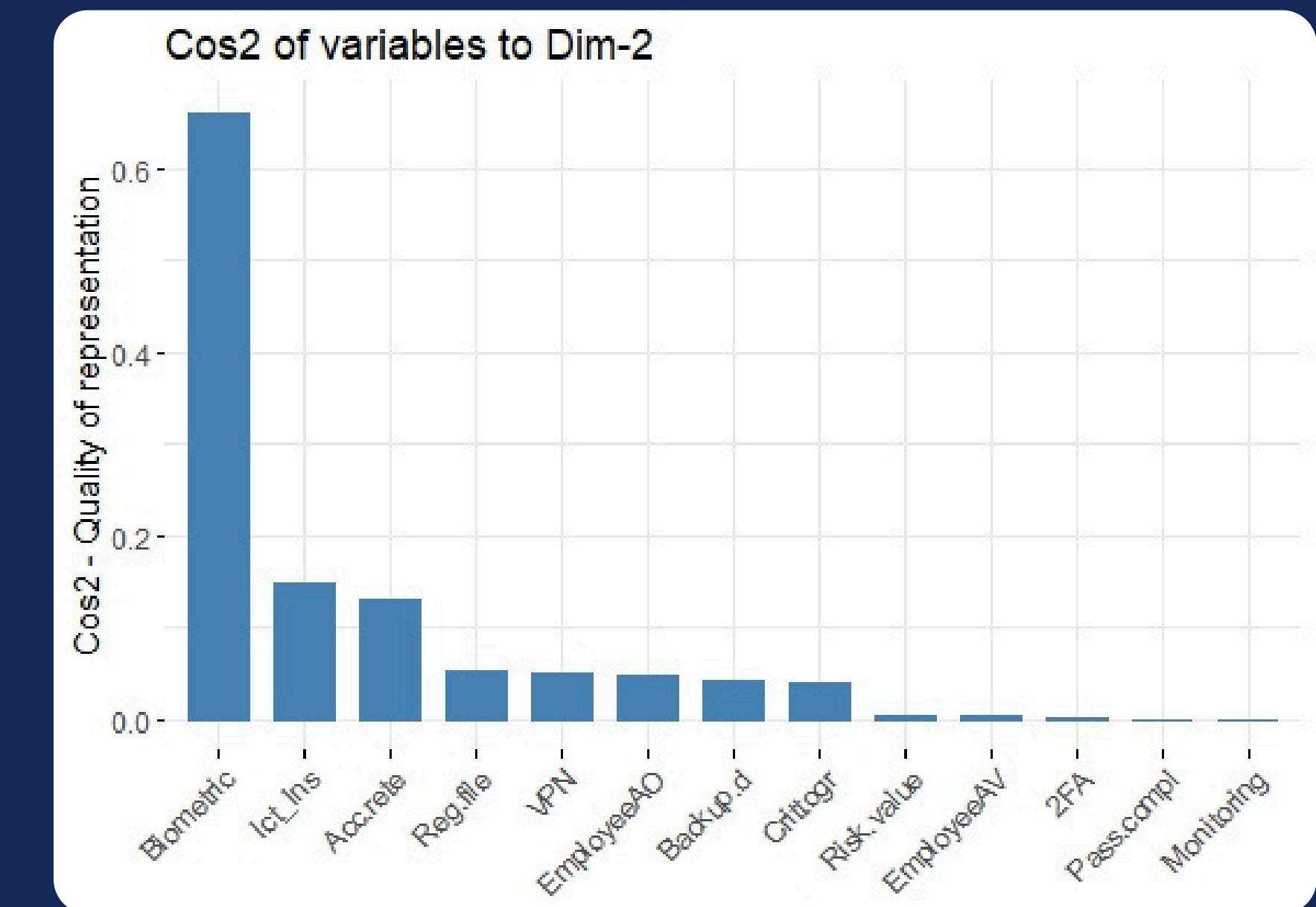
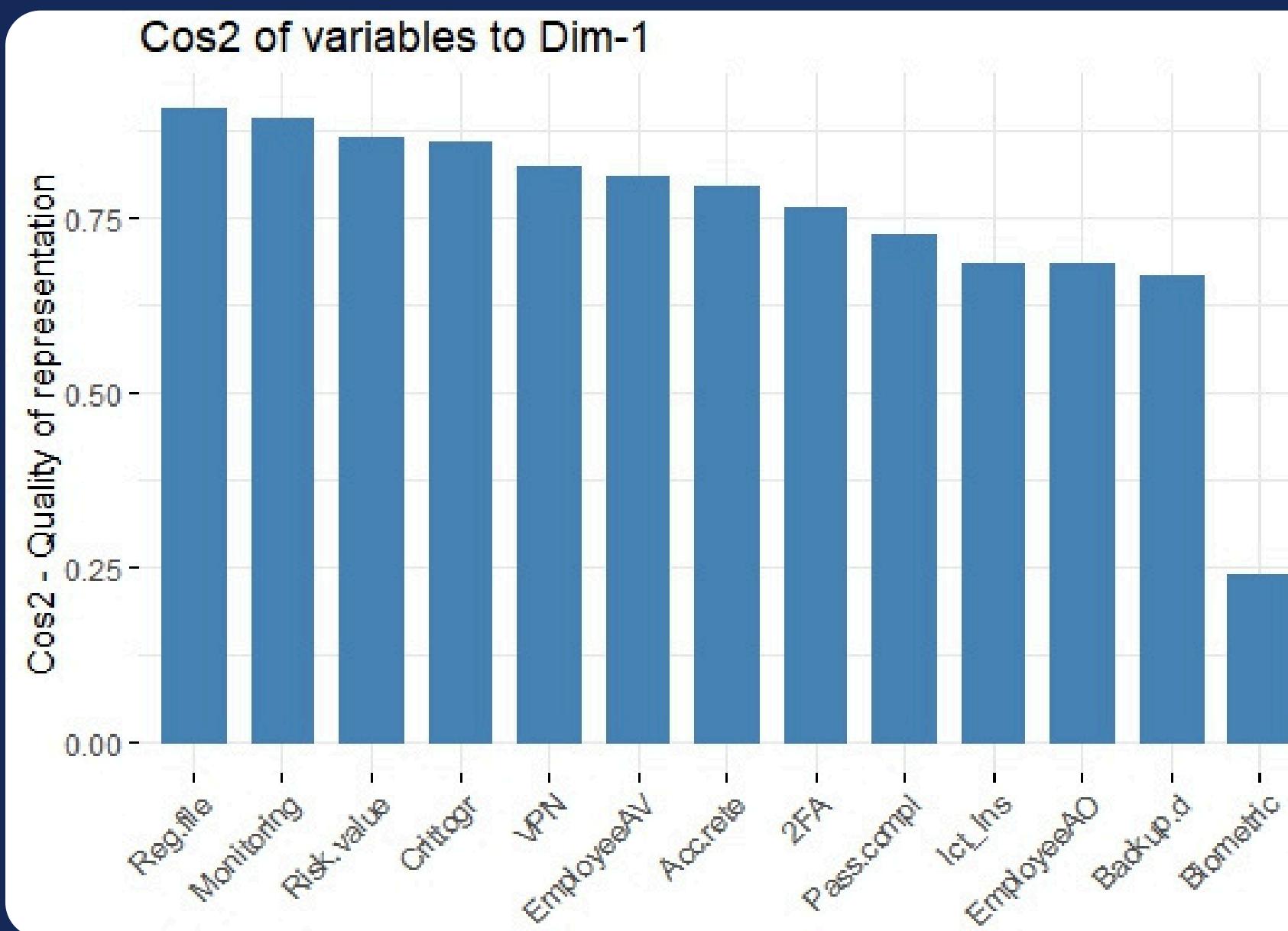


Il grafico dei \cos^2 (coseni quadrati) mostra la qualità della rappresentazione delle variabili nel piano fattoriale dell'ACP.

La maggior parte delle variabili ha una fetta predominante su **Dim.1**, indicando che sono ben rappresentate dalla prima componente, che cattura il comportamento condiviso nei sistemi di sicurezza ICT.

L'unica eccezione significativa è la variabile **Biometric**, rappresentata principalmente da **Dim.2**, confermando il suo ruolo autonomo e la sua specificità rispetto alle altre misure.

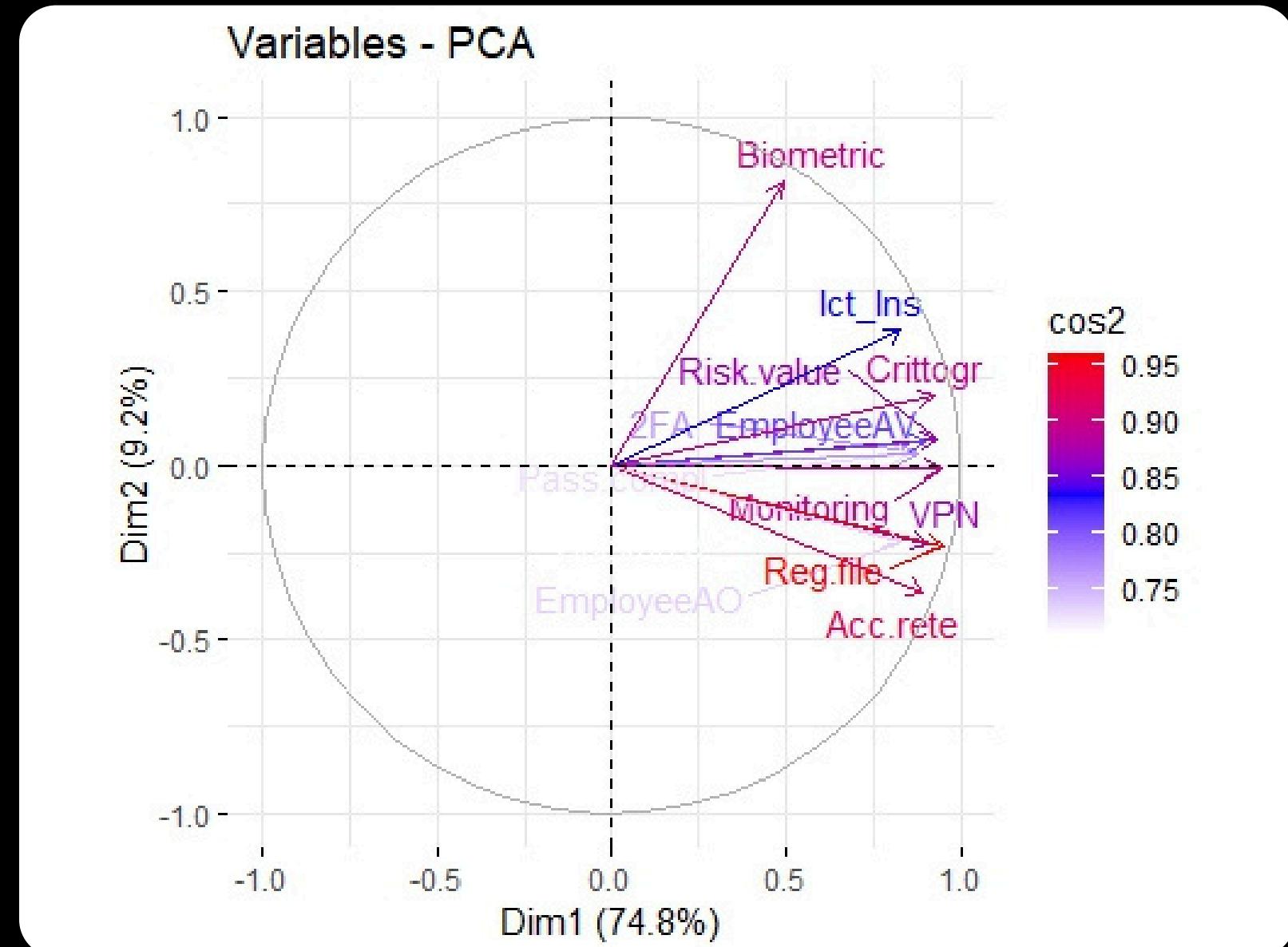
Rappresentazione cos2 delle variabili nelle Dim 1 e Dim 2



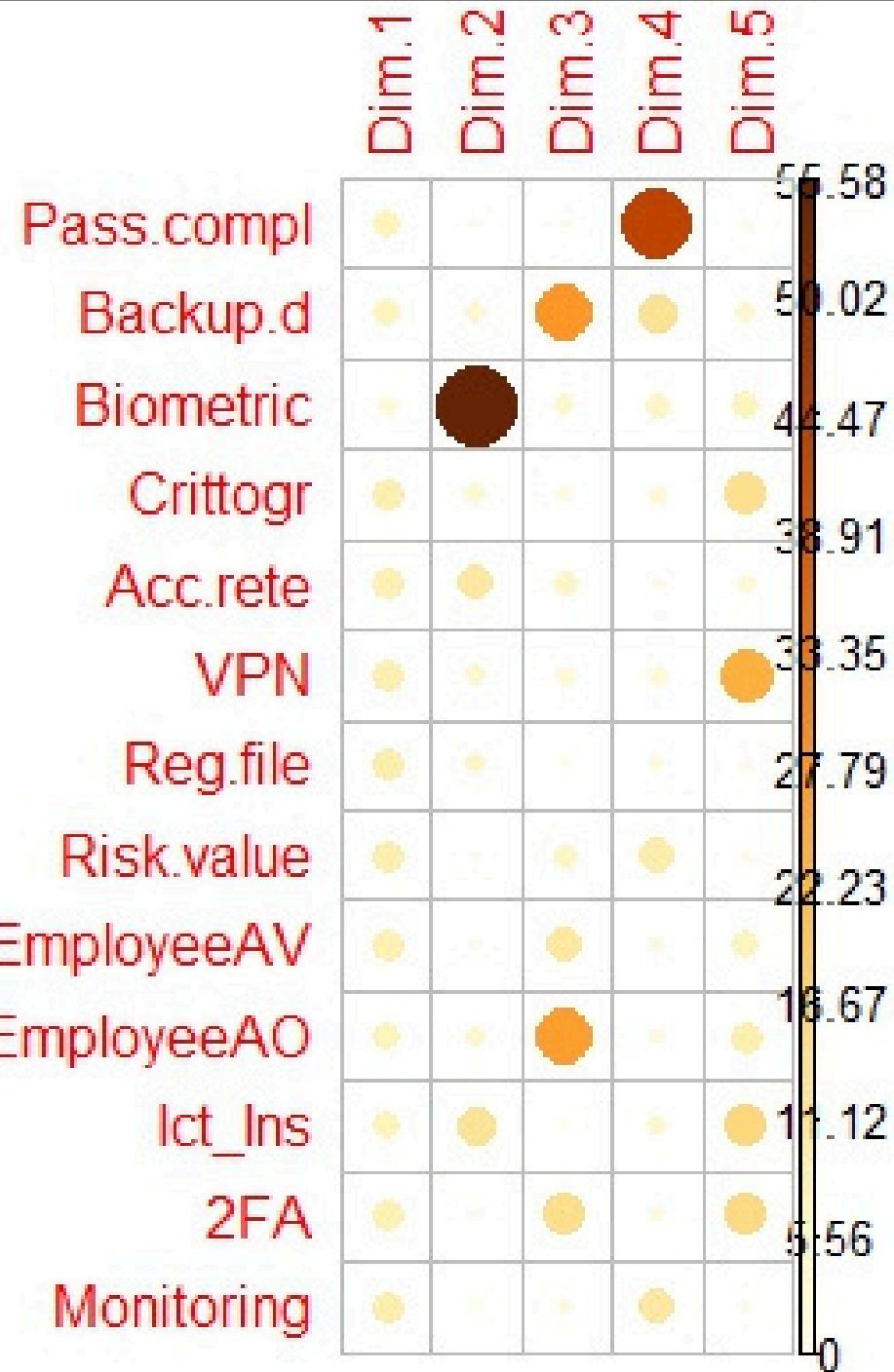
Rappresentazione del cerchio delle correlazioni rispetto ai valori di: Cos2

Le variabili in rosso (es. Reg.file, Monitoring, Acc.rete) e in viola hanno valori di \cos^2 più elevati, e quindi sono molto ben rappresentate nel piano **Dim.1-Dim.2**: queste variabili contribuiscono fortemente alla definizione delle due dimensioni.

La variabile **Biometric** si distingue nettamente dalle altre, mostrando una forte associazione con la seconda componente principale (**Dim.2**)



Contributo delle Variabili:



La prima componente (**Dim.1**) è costruita in modo bilanciato, grazie al contributo simile di quasi tutte le variabili principali.

Questo suggerisce che le imprese italiane non adottano le misure di sicurezza ICT in modo isolato, ma secondo un comportamento coerente e integrato, in cui l'implementazione di una misura è spesso accompagnata da molte altre.

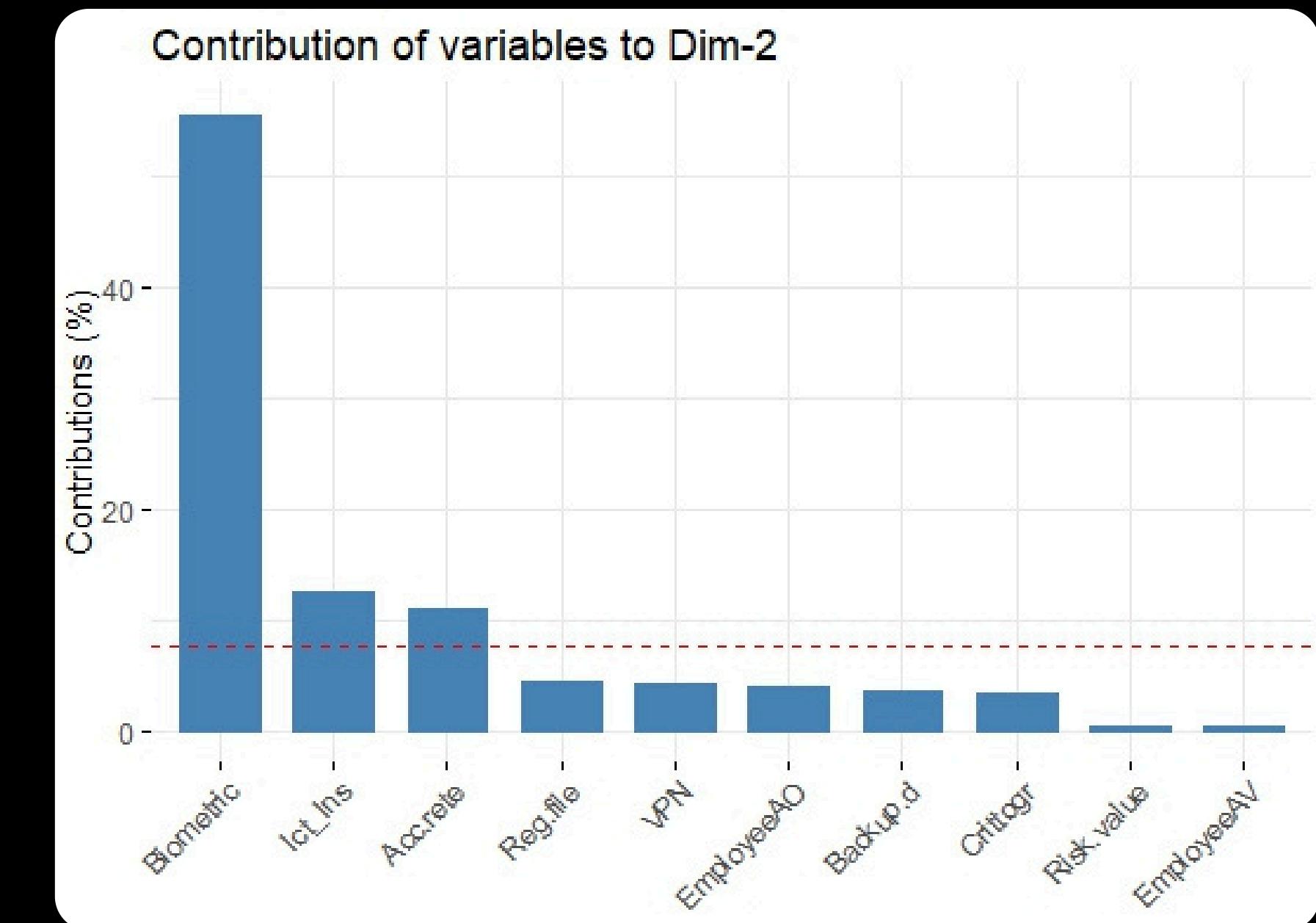
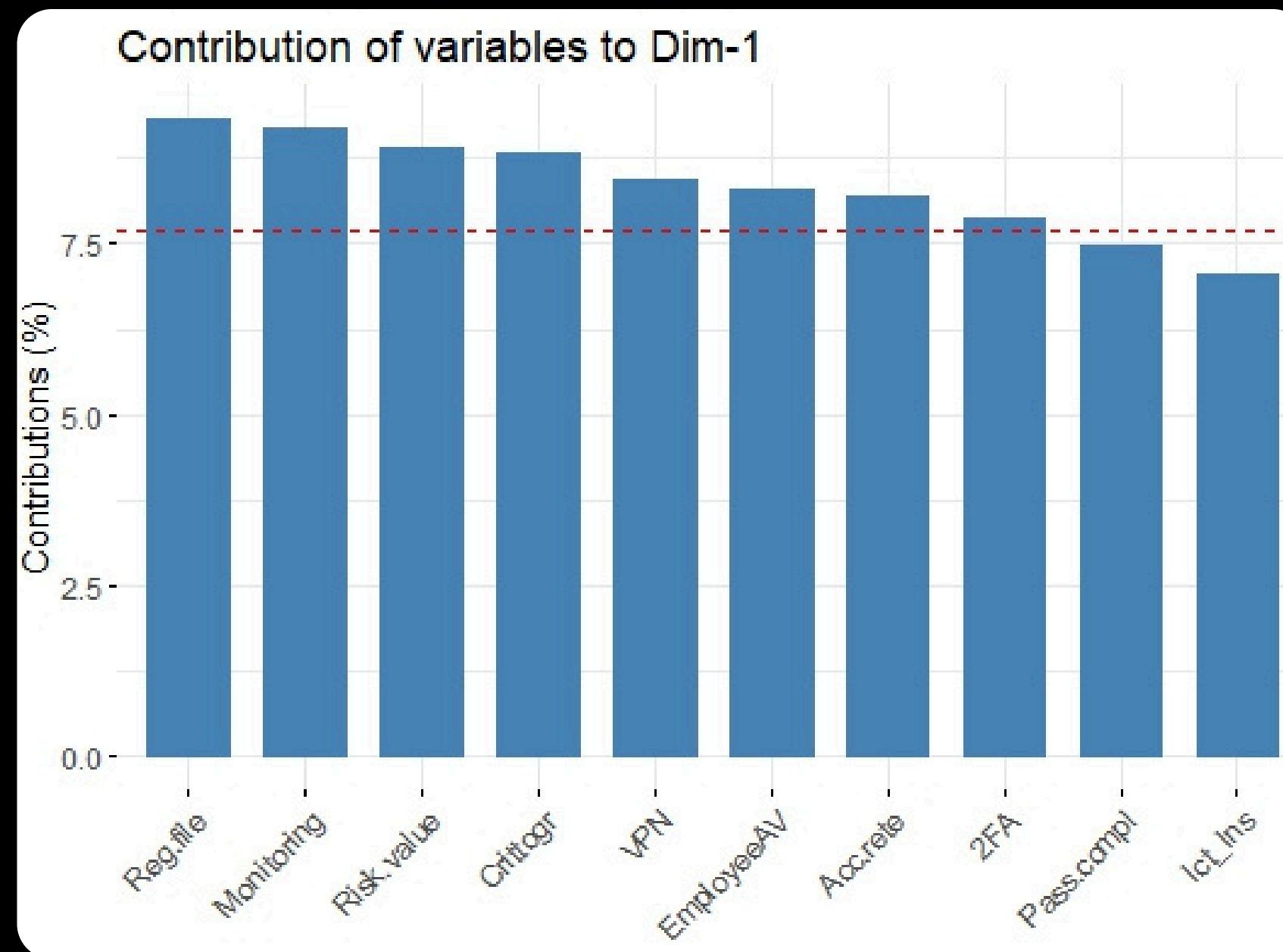
La seconda componente (**Dim.2**) è dominata dalla variabile **Biometric**, che da sola contribuisce in gran parte alla sua costruzione.

Questo suggerisce che **Dim.2** rappresenta una dimensione indipendente, legata all'uso di tecnologie innovative specifiche.

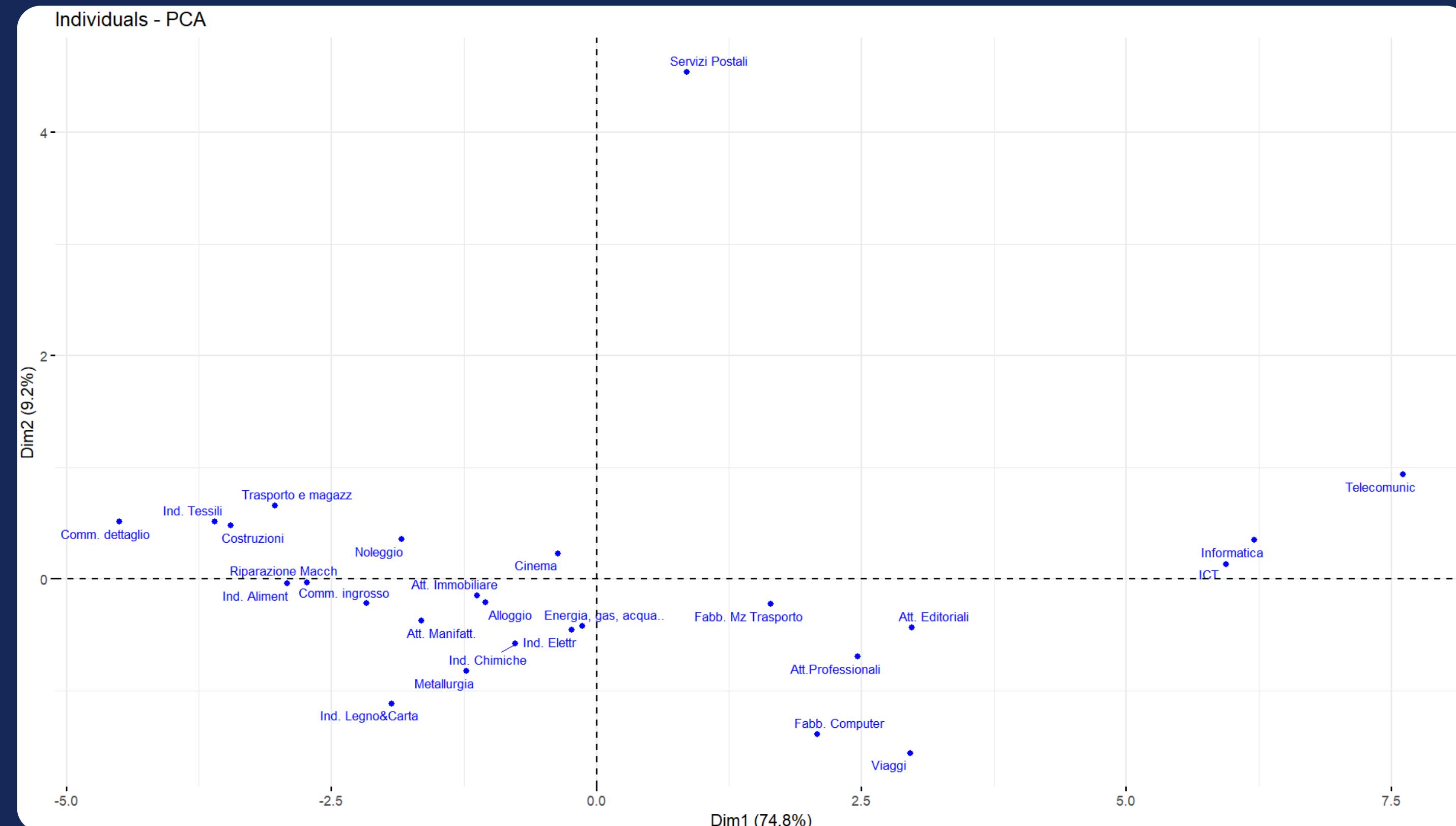
Le componenti successive, pur meno rilevanti a livello interpretativo globale, evidenziano comportamenti specifici di alcune variabili:

- **Dim.3** è influenzata da **Backup.d** e **EmployeeAO**.
- **Dim.4** è influenzata da **Pass.compl** in gran parte.

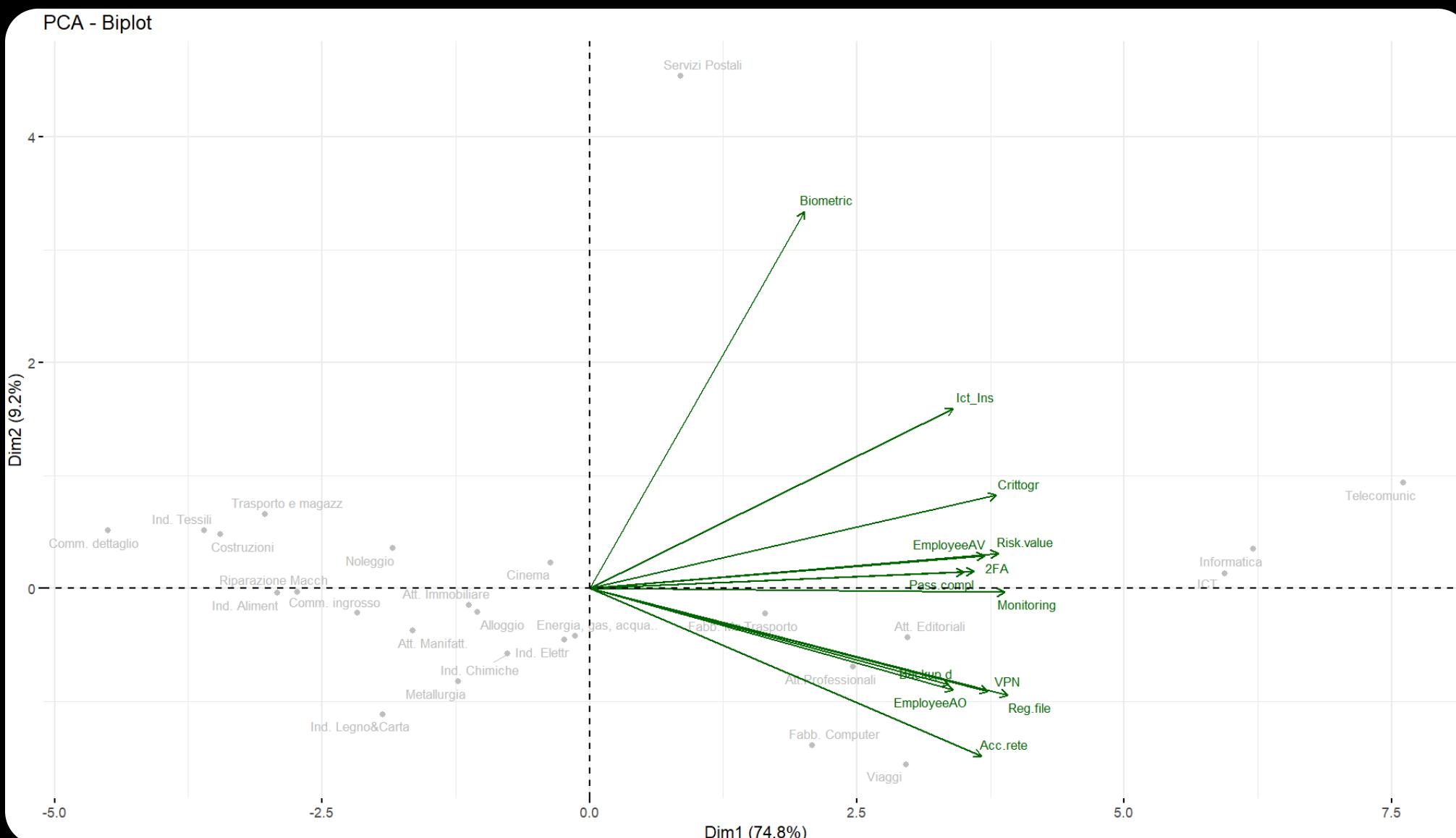
Rappresentazione dei contributi mediante diagramma a barre



Rappresentazione delle Attività su piano fattoriale



Biplot: Variabili x Attività Economiche



Il biplot consente una lettura simultanea delle variabili e delle Attività Economiche (**individui**).

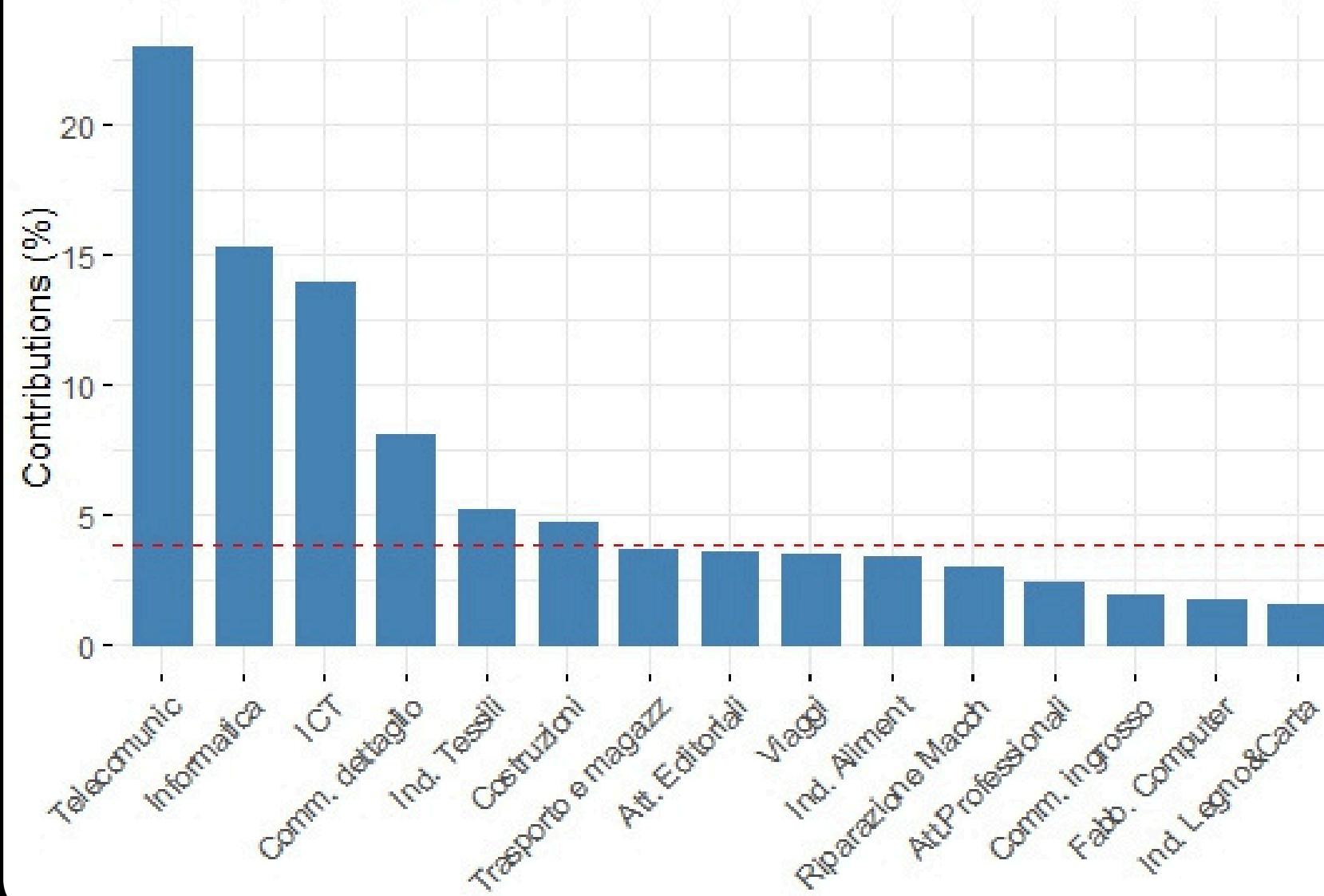
I settori come **Telecomunicazioni, Informatica e ICT** si collocano nella zona dominata dalle variabili **2FA, ICT.test, Reg.file, VPN**, evidenziando una forte adozione di pratiche di sicurezza ICT.

Servizi Postali si distingue per la vicinanza alla variabile **Biometric**, confermando un utilizzo avanzato e indipendente di tecnologie innovative.

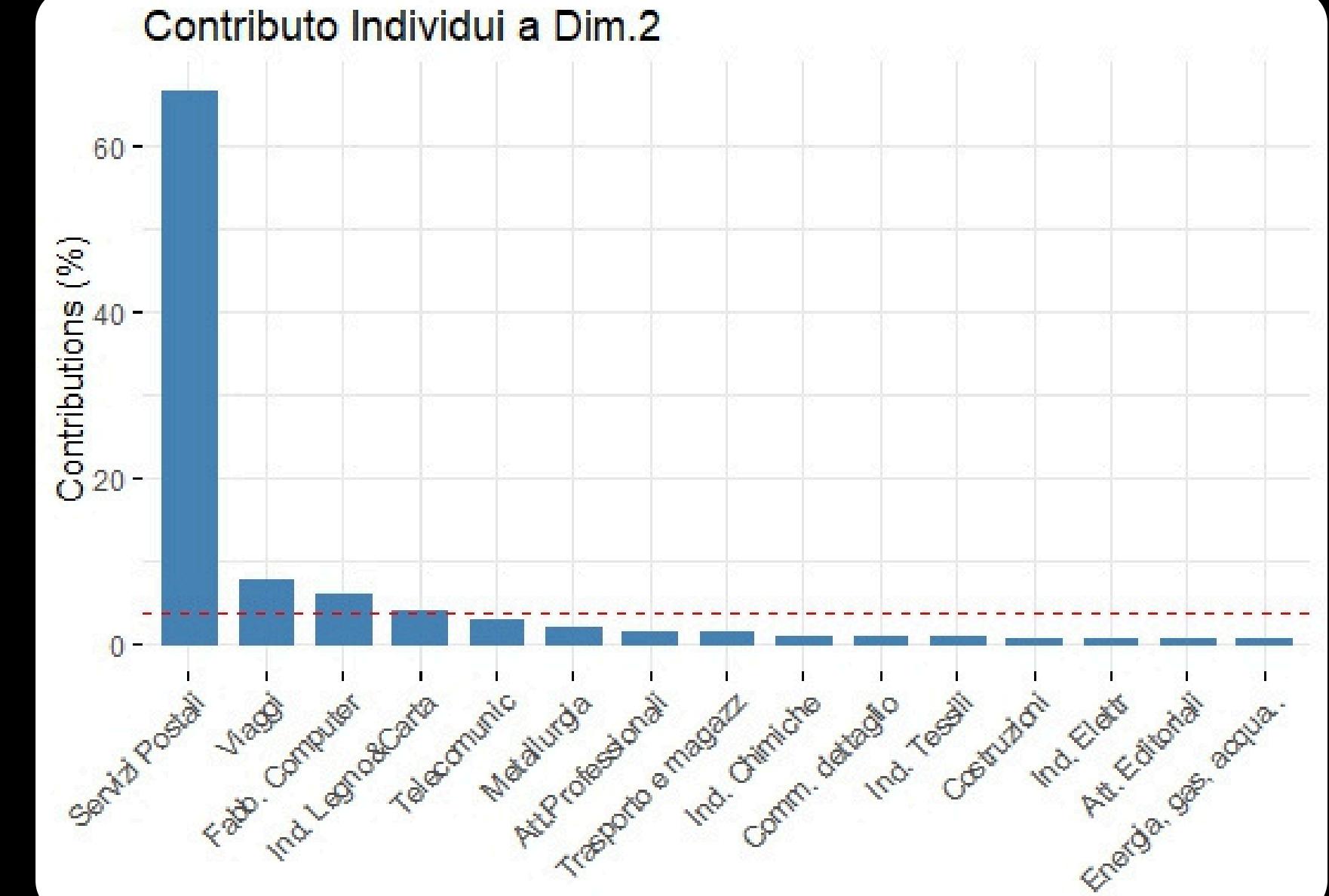
Al contrario, settori come **Commercio al dettaglio, Costruzioni e Industria alimentare** mostrano una distanza significativa dalle variabili di sicurezza, indicando una minore digitalizzazione e un profilo di rischio più elevato.

Contributo delle Attività

Contributo Individui a Dim.1



Contributo Individui a Dim.2



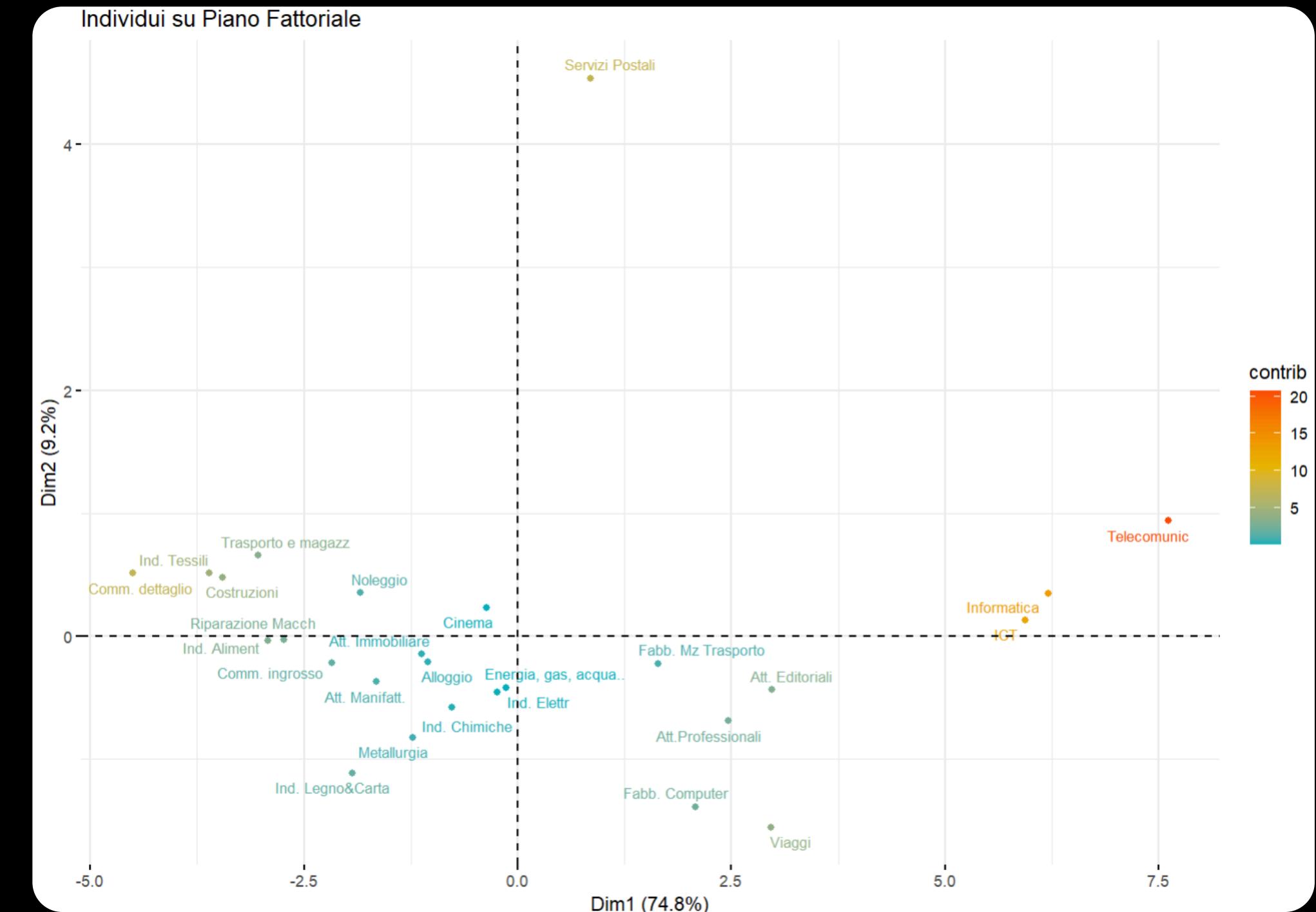
Contributo delle Attività

Il grafico mostra il posizionamento dei settori sul piano principale dell'ACP, con colorazione basata sul contributo di ciascun settore alla costruzione delle componenti.

Alcuni settori, come **Telecomunicazioni** e **Servizi Postali**, emergono per il loro contributo elevato e il posizionamento estremo, indicando strategie ICT molto definite.

Altri settori, pur ben rappresentati (es. ICT, Telecomunicazioni), hanno un ruolo meno dominante.

I settori nella zona centrale, con colori freddi, contribuiscono poco alla sintesi e risultano più omogenei o neutri nella loro adozione di sicurezza ICT.



Cluster Analysis

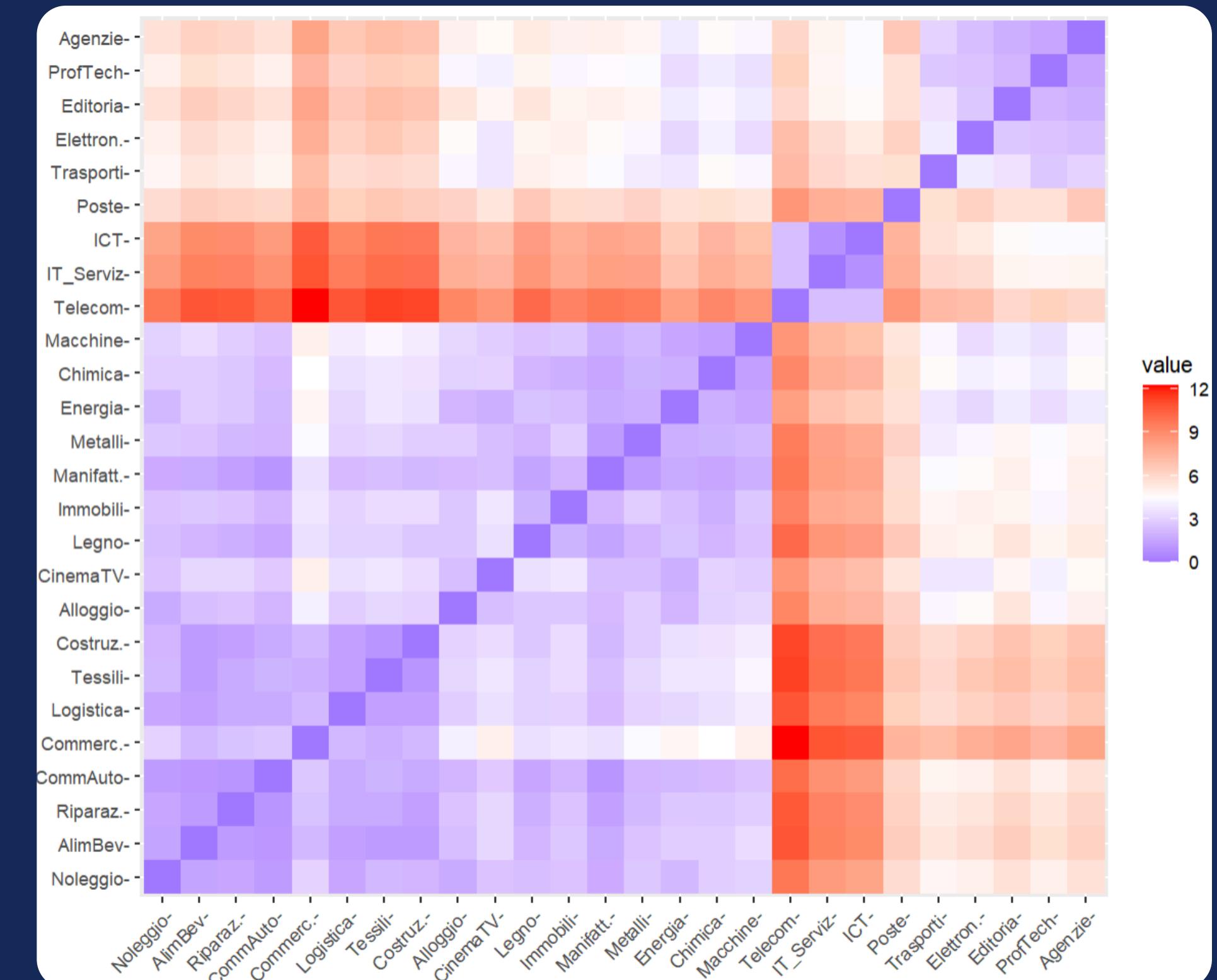


Matrice delle distanze

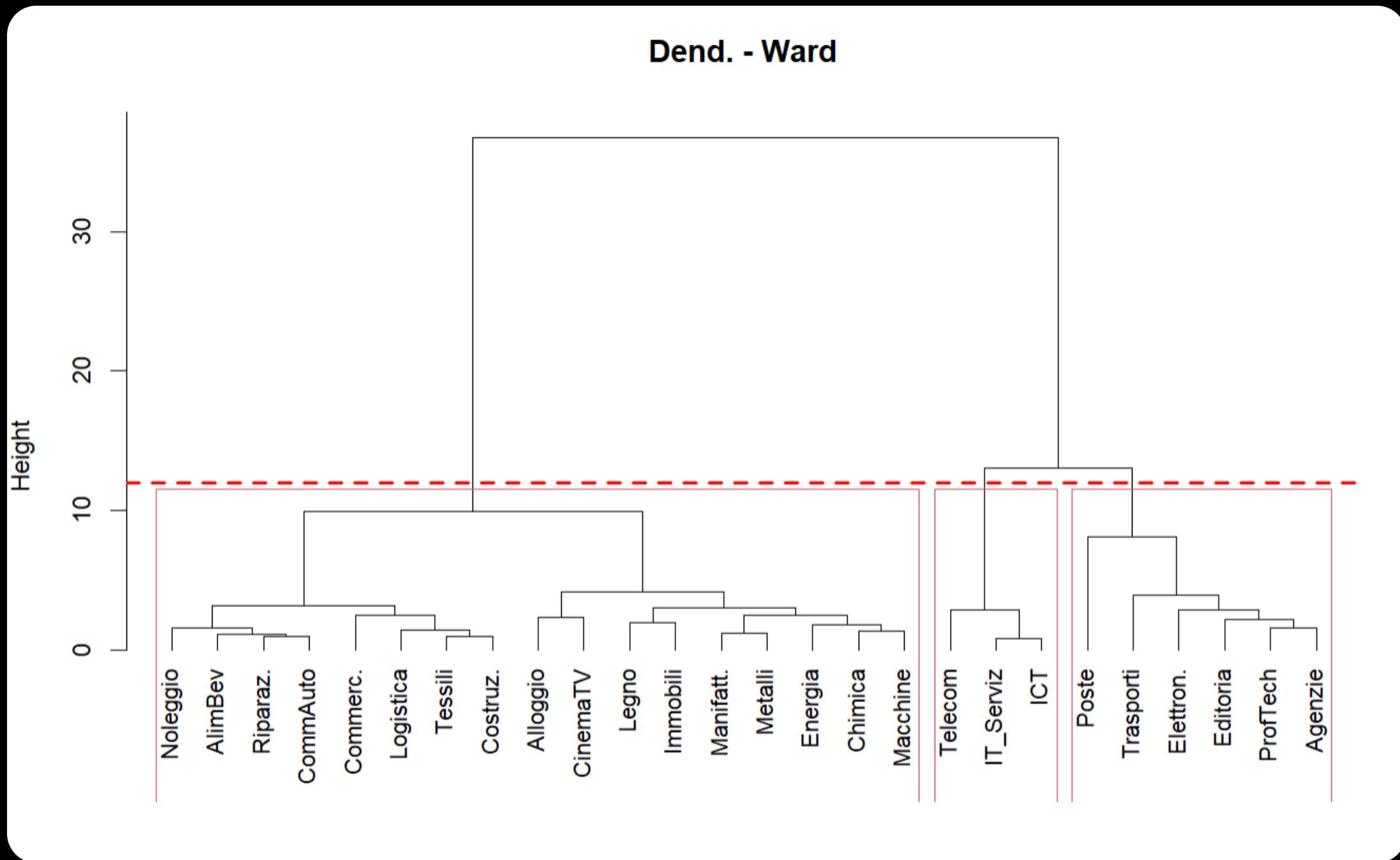
La heatmap mostra la distanza tra attività economiche basata su variabili standardizzate: colori scuri indicano alta similarità, colori chiari maggiore distanza.

Sono evidenti 3 macro-blocchi lungo la diagonale che rappresentano gruppi omogenei:

- **Cluster 1:** Attività di consumo e servizi tradizionali (es. Noleggio, Commercio, Alloggio)
- **Cluster 2:** Attività manifatturiere e industriali (es. Legno, Macchine)
- **Cluster 3:** Attività ad alta intensità tecnologica e professionale (es. Telecomunicazioni, Servizi ICT, Agenzie)



Dendogramma con metodo di Ward



Il dendrogramma mostra come le attività economiche si aggregano in base alla loro similarità, utilizzando il metodo di Ward per minimizzare la varianza interna ai cluster.

Abbiamo scelto di suddividere l'albero ad un'altezza di circa **h=12**

Questo punto ottimale intercetta tre rami principali ben distinti, evitando fusioni forzate tra gruppi già molto dissimili.

Il risultato conferma la struttura osservata nella heatmap delle distanze, evidenziando la presenza di tre cluster naturali nei dati.

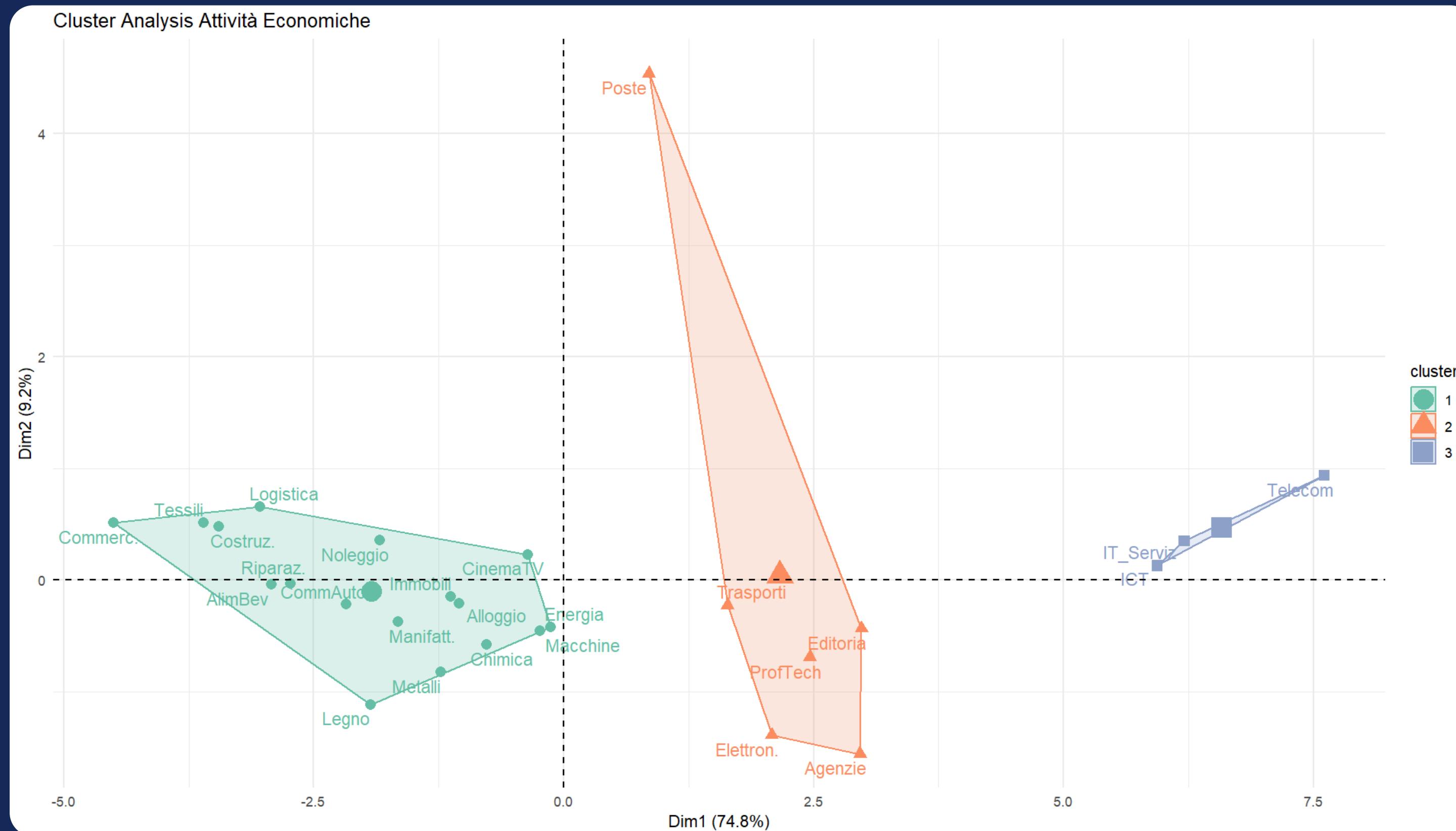
V test

Il V-test evidenzia come i cluster si differenziano rispetto alla media generale:

- **Cluster 1:** mostra valori significativamente più bassi per tutte le misure di sicurezza ICT, indicando una bassa adozione di pratiche di protezione.
- **Cluster 2:** presenta valori moderatamente più alti su alcune variabili (backup, password complesse, accessi sicuri), segnalando un livello intermedio di attenzione.
- **Cluster 3:** si distingue per valori nettamente superiori su quasi tutti gli indicatori di sicurezza, rappresentando contesti con elevata maturità nelle pratiche ICT.

Variabile	v.test	p.value	Cluster
Biometric	-2.65	0.0080	1
EmployeeAO	-2.69	0.0070	1
VPN	-3.51	0.0004	1
Ict_Ins	-3.64	0.0003	1
Crittogr	-3.77	0.0002	1
Acc.rete	-3.78	0.0002	1
2FA	-3.79	0.0001	1
Reg.file	-3.86	0.0001	1
Backup.d	-3.88	0.0001	1
Pass.compl	-3.92	0.0001	1
Monitoring	-3.97	0.0001	1
EmployeeAV	-3.98	0.0001	1
Risk.value	-4.11	0.0000	1
Backup.d	2.95	0.0031	2
Pass.compl	2.33	0.0196	2
Acc.rete	2.32	0.0205	2
Risk.value	2.12	0.0337	2
EmployeeAO	4.15	0.0000	3
Crittogr	4.04	0.0001	3
2FA	3.91	0.0001	3
EmployeeAV	3.87	0.0001	3
Ict_Ins	3.68	0.0002	3
Reg.file	3.50	0.0005	3
Monitoring	3.44	0.0006	3
VPN	3.37	0.0007	3
Risk.value	3.33	0.0009	3
Pass.compl	2.75	0.0059	3
Acc.rete	2.58	0.0100	3
Biometric	2.13	0.0329	3

HCPC: Hierarchical Clustering on Principal Components



Terza analisi

Text Mining



In questa ultima sezione, sono state applicate tecniche di **text mining** su un dataset contenente descrizioni di incidenti informatici avvenuti dal 2005 al 2020. Il dataset è stato costruito su incidenti raccolti dal Council on Foreign Relations (CFR), una think tank statunitense specializzato in politica estera e affari internazionali.



L'obiettivo principale è stato identificare i temi ricorrenti negli incidenti informatici, al fine di comprendere le minacce prevalenti e le tendenze emergenti nel settore della cybersecurity. Il lavoro è stato articolato in sei fasi

Metadati principali

- **Fonte:** Kaggle/CFR - Cyber incidents from 2005 to 2020
- **Anno di riferimento:** 2005-2020
- **Unità di analisi:** Descrizione degli incidenti avvenuti
- **Formato:** CSV
- **Copertura geografica:** Mondiale
- **Tipo di dati:** testuali

Pre-trattamento

- Filtraggio della variabile utile all'indagine (*Description*)
- Rimozione dei dati mancanti

Title	Date	Affiliations	Description	Response	Victims	Sponsor	Type	Category	Sources_1	Sources_2
Attack on . 1/2/2020 Turla			The suspect Confirms	Austrian Fed	Russian Fed	Espionage	Governme	https://www.h	https://www.h	https://www.h
Spear-phis 1/23/2020 Konni Gro		The suspected North	Korean Group	Employees	Korea (Der	Espionage	Governme	https://unit42.p	https://unit42.p	https://unit42.p
Australian 4/6/2020			Responsible for attacking	infrast	Australia	Data destr	Private sec	https://ww	https://ww	https://ww
Catfishing 2/16/2020 APT-C-23		The Hama	Hack Back	Israeli Def	Palestine,	Espionage	Military	https://ww	https://ww	https://ww
Targeting 4/8/2020 Fox Kitten		Iranian hackers attac	U.S. gover	Iran (Islam	Espionage	Governme	https://ww	https://ww	https://ww	https://ww
Tracking o 3/29/2020 Governor		The Saudi Arabian go	Saudi Arab	Saudi Arab	Espionage	Private sec	https://ww	https://ww	https://ww	https://ww
Targeting o 5/12/2020 Hidden Co		Hidden Co Denounce	Global fina	Korea (Der	Financial T	Private sec	https://ww	https://ww	https://ww	https://ww
Pioneer Kit 8/31/2020		Also known	Responsible for gathe	Iranian exp	Iran (Islam	Espionage	Civil societ	https://www.cro	https://www.cro	https://www.cro
Targeting o 1/13/2020 APT 28		The alleged Russian a	Burisma	Russian Fed	Espionage	Private sec	https://ww	https://ww	https://ww	https://ww
Targeting o 1/28/2020 KINGDOM		The Saudi hackers KIN	Ben Hubba	Saudi Arab	Espionage	Civil societ	https://citizenlab	https://citizenlab	https://citizenlab	https://citizenlab
Targeting o 7/17/2020 Believed t		Hackers ta Confirmat	Israel	Iran (Islam	Sabotage	Governme	https://ww	https://ww	https://ww	https://ww
Targeting o 4/15/2020 Syrian Elec		The Syrian Electronic	Arabic spe	Syrian Aral	Espionage	Civil societ	https://ww	https://ww	https://ww	https://ww
Targeting o 8/14/2020 Lazarus Gr		Responsible for using	Defense co	Korea (Der	Espionage	Governme	https://www.cle	https://www.cle	https://www.cle	https://www.cle
Targeting o 2/16/2020 APT 34		The Iranian threat act	Private co	Iran (Islam	Espionage	Private sec	https://www.cle	https://www.cle	https://www.cle	https://www.cle
Targeting o 4/23/2020 Russia		Hackers br Denounce	Poland's W	Russian Fed	Defaceme	Governme	https://ww	https://ww	https://ww	https://ww
Targeting o 9/18/2020 Pioneer Kit		Iranian hackers stole	Iranian exp	Iran (Islam	Espionage	Civil societ	https://res	https://res	https://res	https://res
Compromi 1/8/2020	APT 34	The Iranian state-bac	Bapco	Iran (Islam	Espionage	Private sec	https://ww	https://ww	https://ww	https://ww
Storm Clou 3/31/2020	NOTE: Son	Performed highly targeted water	China							
Targeting o 7/28/2020 RedDelta		Hackers cc Unknown	Vatican Cit	China						
Targeting o 3/5/2020	Tonto Tea	The Chinese governm	Japanese, Ch							
Operation: 4/6/2020	Australian	The Australian govern	Infrastruct Austral							
BlackTech 8/19/2020		Responsible for targe	Taiwanese Ch							
Targeting o 1/9/2020	APT 33	The Iranian hackers A	The U.S. el	Iran (Is						
Targeting o 6/17/2020	Believed t	Hackers impersonate	European	Korea (Der						
Targeting o 11/14/2020	Believed t	Chinese ha Confirmat	U.S. gover	China						
Stolen dat 1/20/2020	Bronze Bu	The China-linked Bror	Mitsubishi	China						
Targeting o 3/31/2020	Storm Clou	The Chinese hackers S	Tibetans	China						
Targeting o 7/21/2020	Guangdon	Hackers cc Criminal cl	Private Co	China						

▲ Description

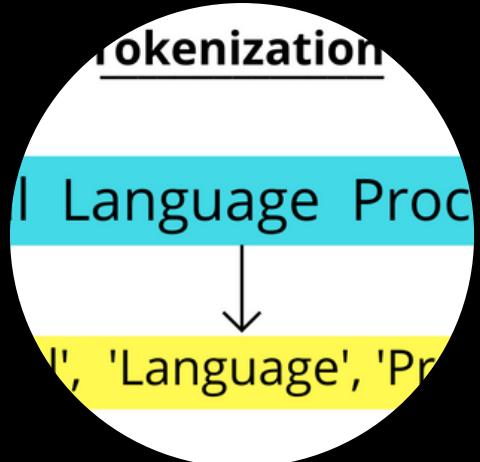
- 1 The suspected Russian hackers conducted a weeks-long att...
- 2 The suspected North Korean threat actor Konni Group atte...
- 3 Responsible for attacking infrastructure that cyber criminals ...
- 4 The Hamas-associated threat actor APT-C-23 targeted Israel...
- 5 Iranian hackers attacked high-end networking equipment wi...
- 6 The Saudi Arabian government used flaws in the global Sign...
- 7 Hidden Cobra used a variety of malware tools to hack into a...
- 8 Responsible for gathering sensitive and private information ...
- 9 The alleged Russian actors hacked Burisma, a Ukrainian gas ...
- 10 The Saudi hackers KINGDOM launched a phishing attack to ...
- 11 Hackers targeted two water pumps in Galilee and Mateh Ye...
- 12 The Syrian Electronic Army distributed numerous COVID-19...
- 13 Responsible for using social engineering and lofty job posti...
- 14 The Iranian threat actors targeted dozens of companies in t...
- 15 Hackers breached Poland's War Studies University's website ...

Fasi del lavoro



Fase 1: Importazione e pulizia

- Importazione del dataset pre-trattato
- Creazione di una funzione per pulire la variabile description
 - Trasformazione dal maiuscolo al minuscolo
 - Rimozione punteggiatura e numeri
 - Rimozione stopwords inglesi e di eventuali spazi bianchi



Fase 2: Tokenizzazione

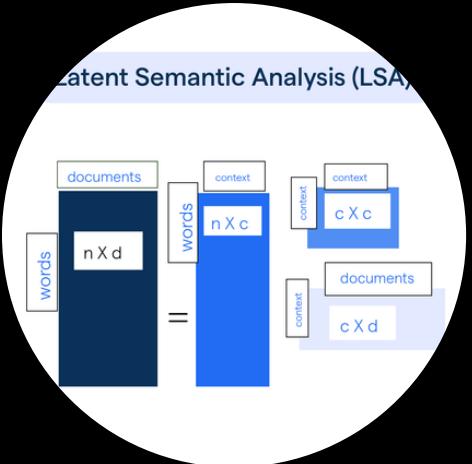
- Creazione dei nomi univoci per i documenti
- Creazione del corpus per tokenizzare, contenente le descrizioni degli incidenti informatici
- Tokenizzazione del testo per suddividerlo in unità più piccole
- Rimozione di stopwords inutili e generiche per evitare rumore

Fasi del lavoro

	words1
doc1	0
doc2	2
doc3	0
doc4	0

Fase 3: Creazione della DFM

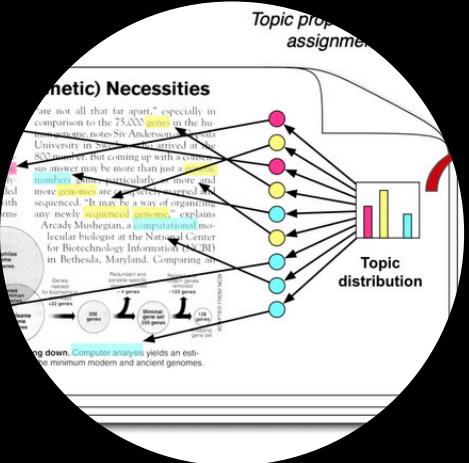
- Creazione di una matrice **documento-feature** (nel nostro caso più specifica e performante rispetto alla DocumentTermMatrix)
- Rimozione di parole rare che compaiono in meno di cinque documenti



Fase 4: Latent Semantic Analysis

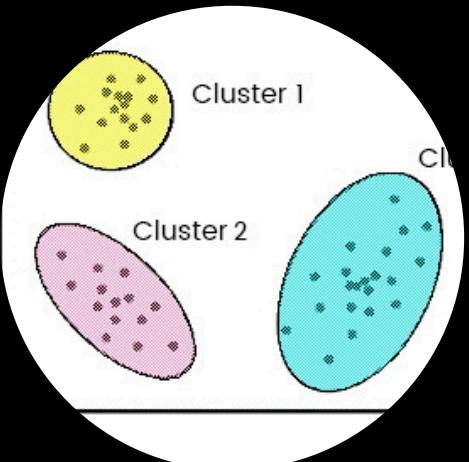
- Applicazione della tecnica di riduzione dimensionale tramite Singular Value Decomposition (**SVD**)
- Riduzione del numero di dimensioni per focalizzarsi sui principali concetti latenti
- Creazione di una matrice ridotta con i principali “topic” dei dati testuali

Fasi del lavoro



Fase 5: Topic Modelling

- Applicazione del modello **LDA** (Latent Dirichlet Allocation) per identificare i topic latenti nei dati
- Estrazione delle parole più rappresentative per ciascun topic con i relativi beta (term probability)
- Visualizzazione attraverso un ortogramma orizzontale

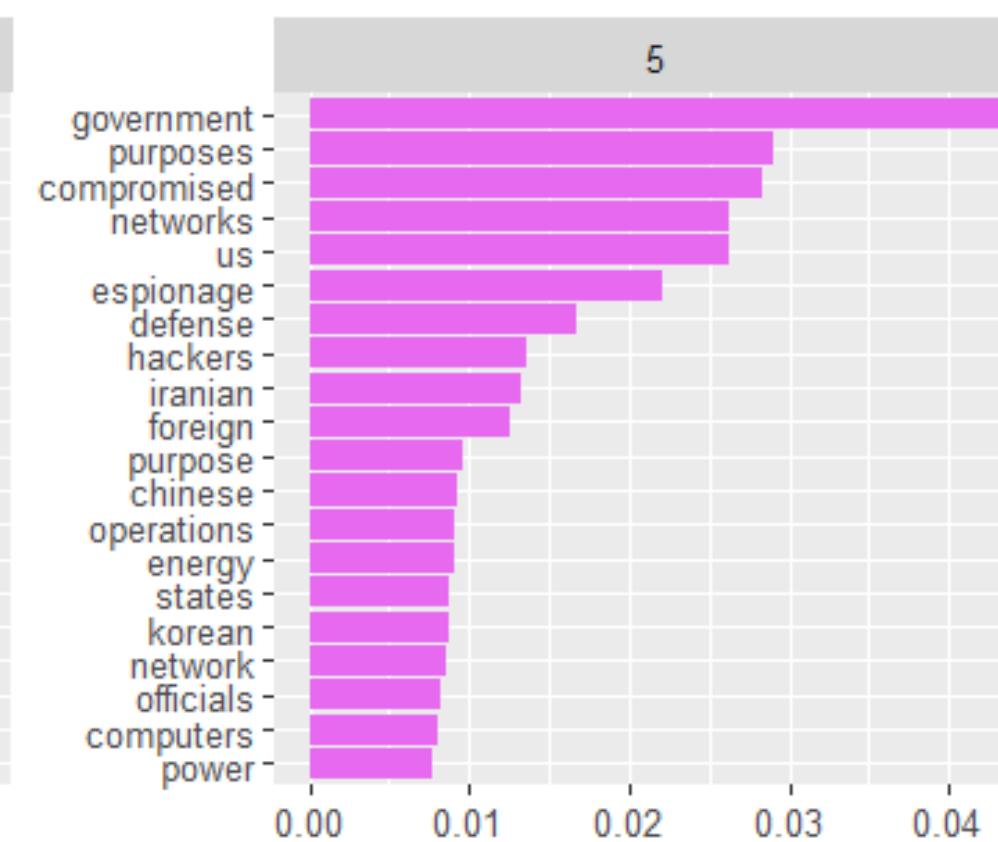
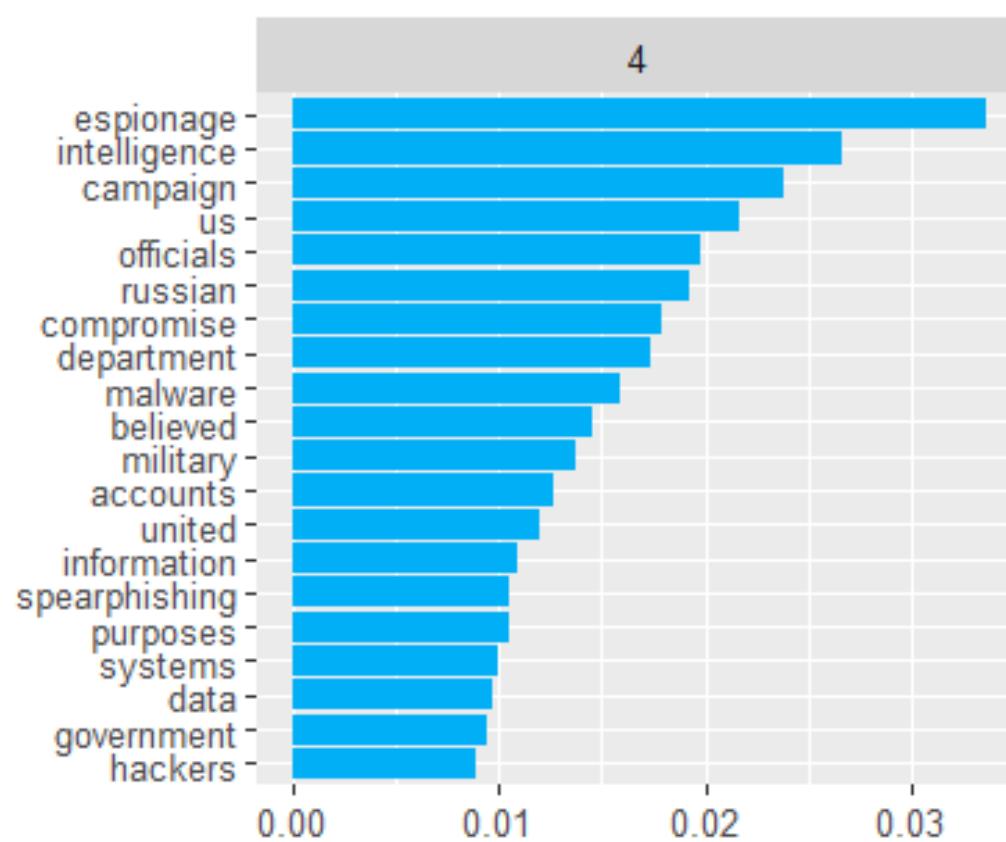
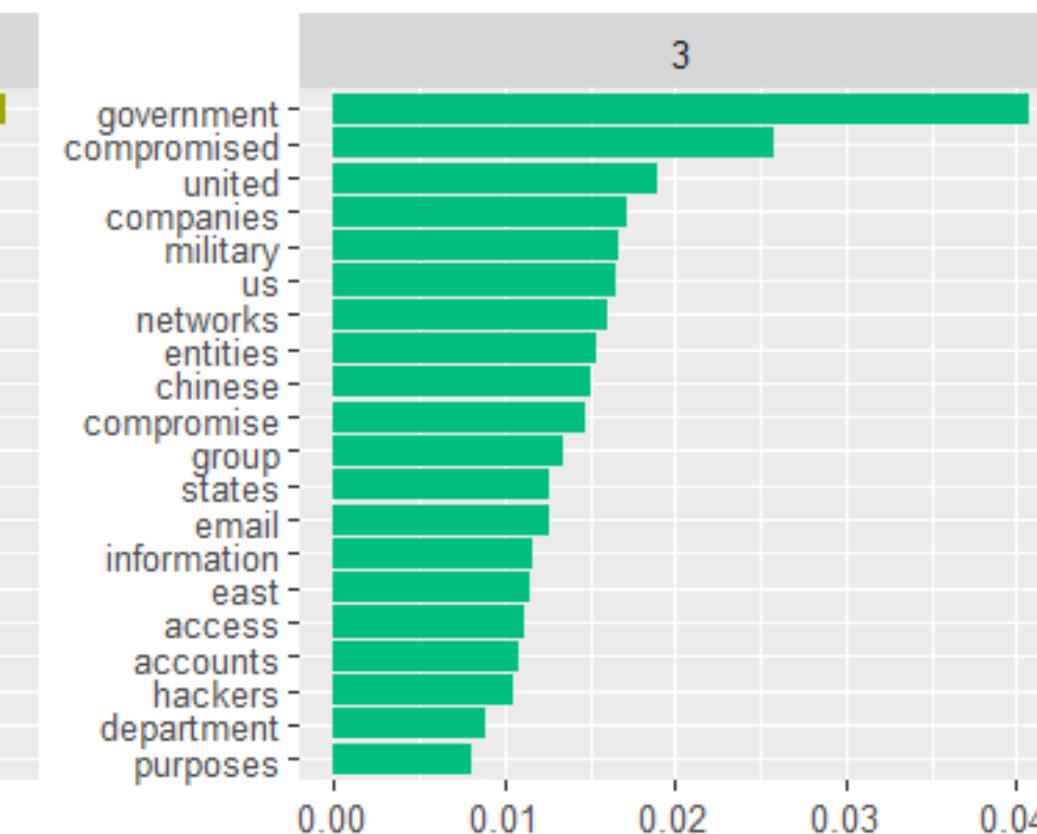
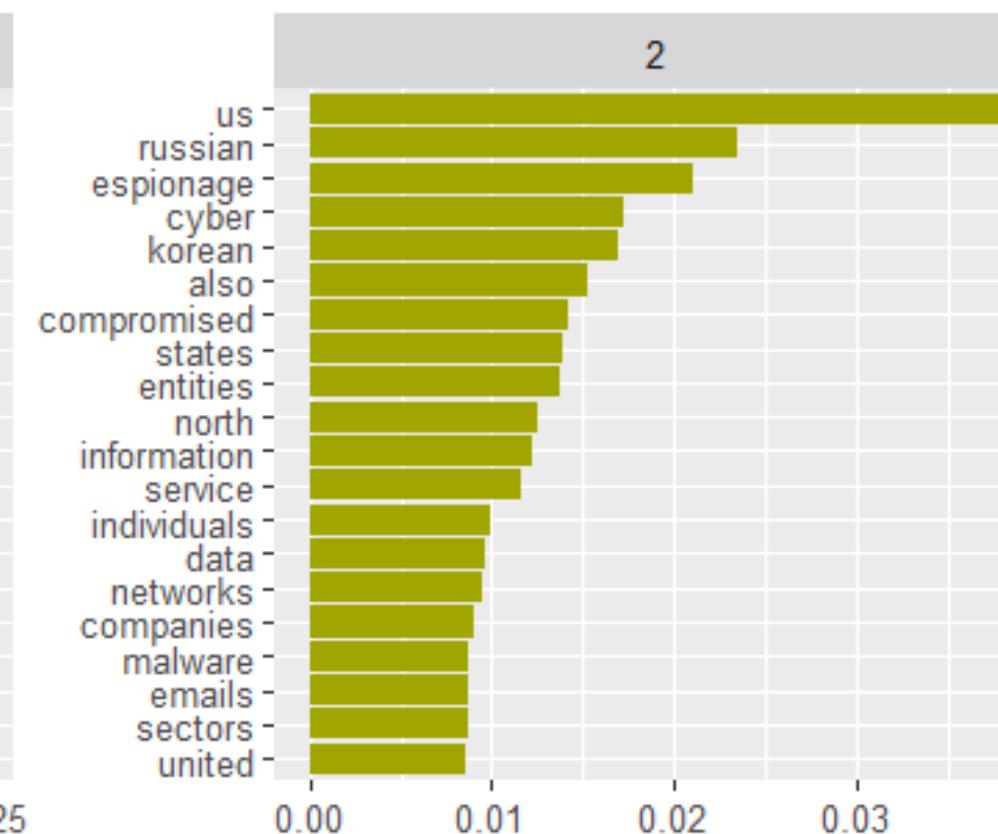
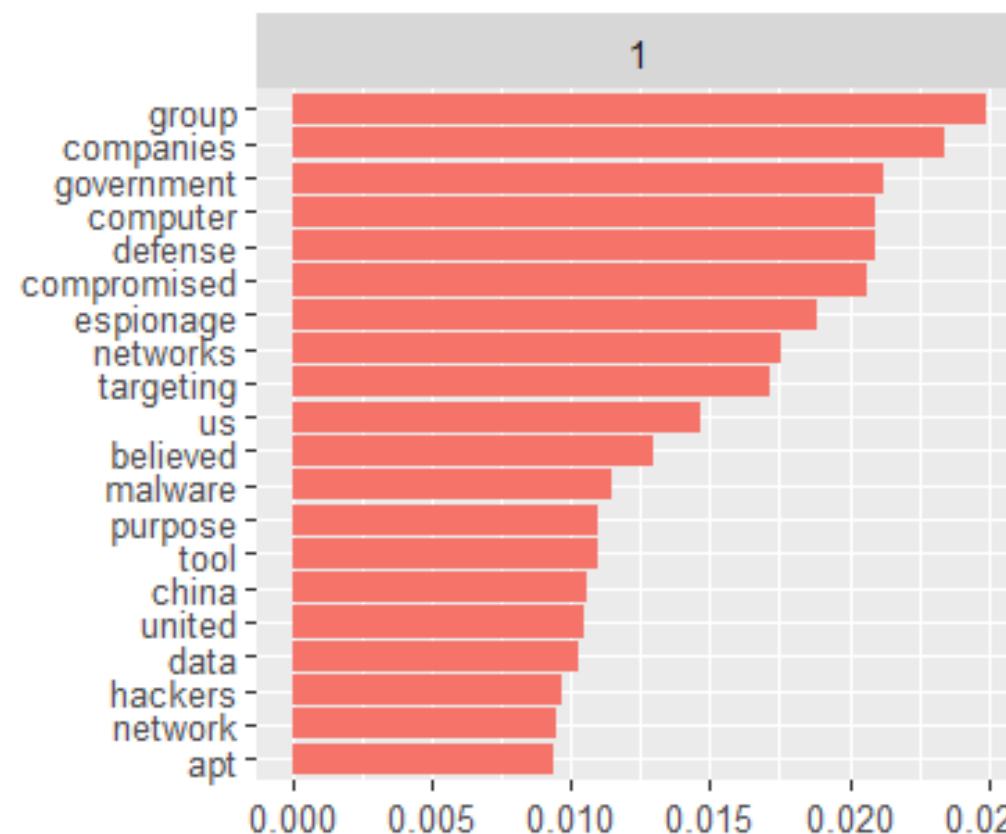


Fase 6: Cluster Analysis

- Applicazione del clustering **K-means** per suddividere in gruppi (cluster)
- Esplorazione qualitativa dei cluster, mostrando alcune descrizioni esemplificative per ogni gruppo

Distribuzione dei topic

Top 20 Termsper Topic



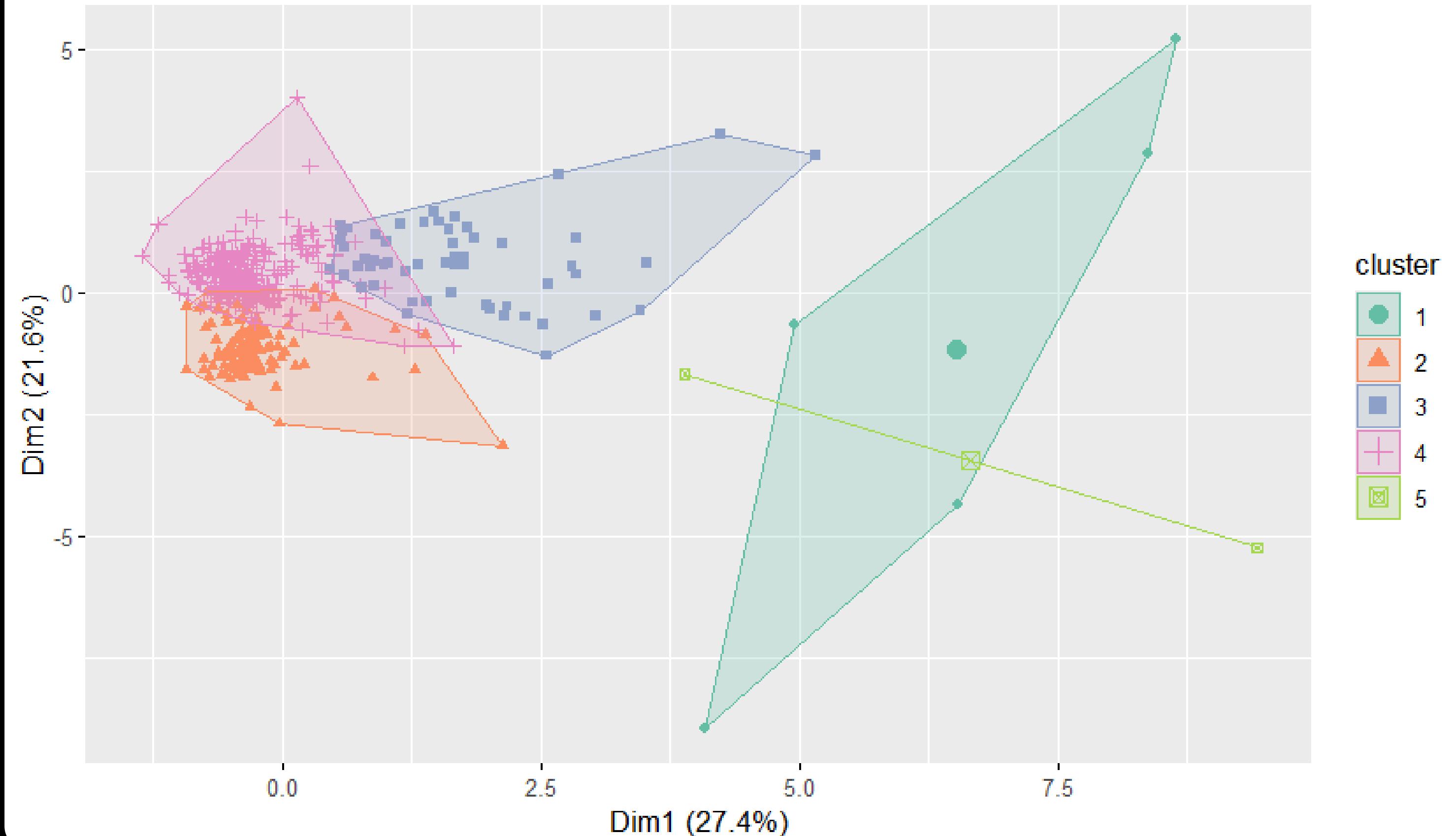
Beta (Term Probability)

Risultati

Topic	Tema Principale	Attori Coinvolti	Obiettivi	Tecniche Usate	Keywords
1	Attacchi APT a enti governativi e aziende	Gruppi hacker (es. APT), Cina	Reti aziendali, governi, infrastrutture critiche	Spionaggio, malware, targeting reti	group, companies, defense, espionage, china, APT
2	Cyber-espionaggio da Russia e Corea del Nord	Russia, Corea del Nord	Stati Uniti (governo, settori strategici)	Compromissione dati, malware, attacchi statali	russian, korean, espionage, entities, emails
3	Infiltrazioni in infrastrutture USA	Cina, gruppi hacker	Governi, aziende, forze armate USA	Furto account/email, compromissione reti	compromised, military, chinese, east, email
4	Tattiche avanzate di spionaggio informatico	Russia, agenzie di intelligence	Governi, ufficiali militari	Spearphishing, malware, account compromise	espionage, intelligence, spearphishing, officials
5	Minacce da nazioni straniere (Iran, Cina, ecc.)	Iran, Cina, Corea del Nord, attori esteri	Difesa, energia, infrastrutture critiche	Spionaggio, compromissione reti, attacchi mirati	iranian, power, network, foreign, energy, defense

Cluster Analysis

Cluster Analysis degli Incidenti Cyber (LSA + KMeans)



Risultati

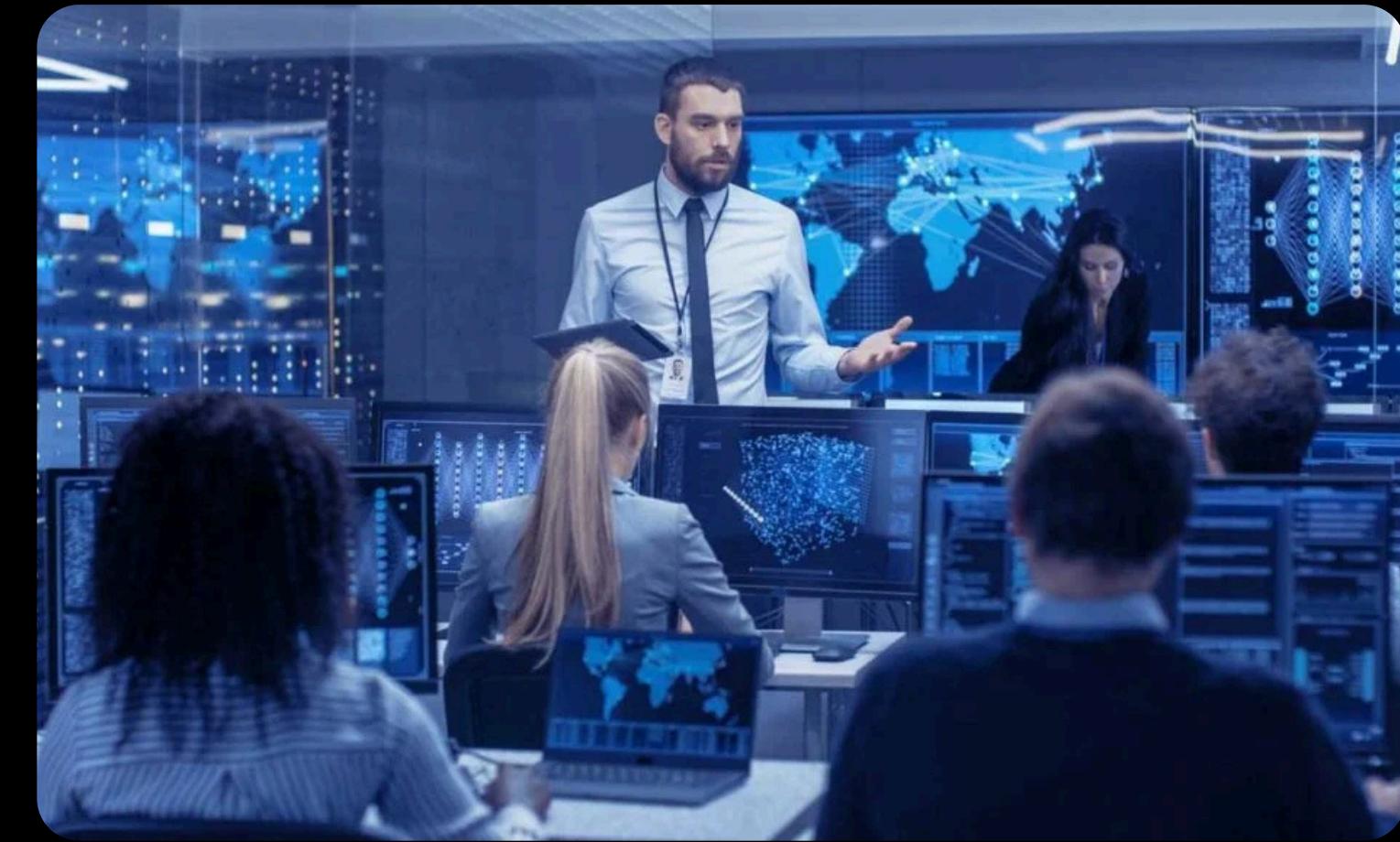
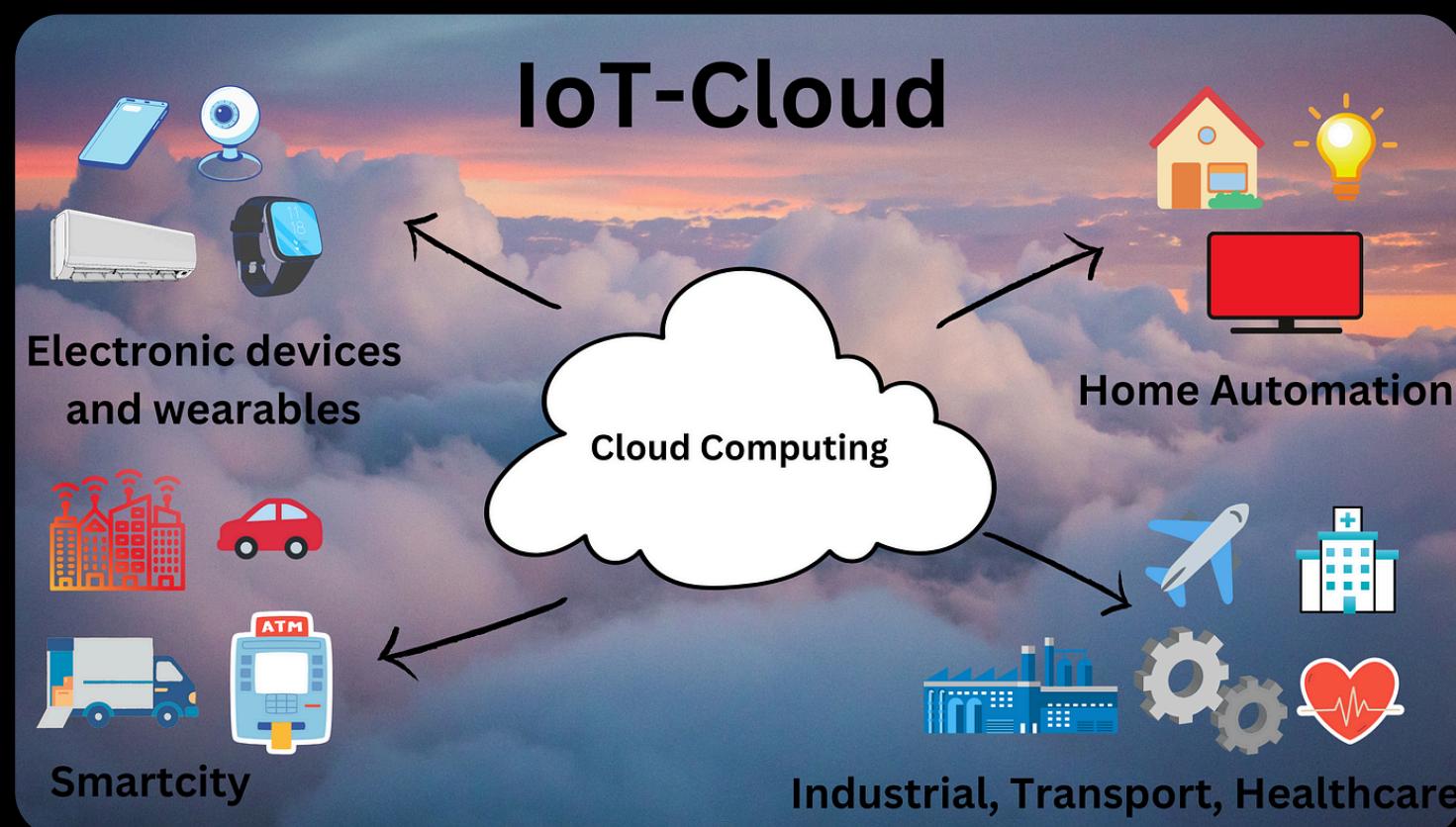
Cluster	N. Documenti	Contenuto Principale	Temi Chiave	Topic LDA Correlati
1	1	Caso complesso (GRU, hacking militare ucraino, media europei)	Spionaggio statale, attacchi sofisticati, geopolitica	Topic 1 e 2
2	34	Cyber Command USA vs GRU, sabotaggi, interferenze elettorali, Westinghouse, FIFA, università	Guerra cibernetica, interferenze politiche, controspionaggio	Topic 2 e 4
3	117	Cluster più ampio: sabotaggi, giornalisti, infrastrutture, Corea, Iran, Hamas, Cina, social engineering, malware	Spionaggio, disinformazione, manipolazione geopolitica	Topic 2 e 4
4	327	Campagne APT in Asia e Medio Oriente, targeting di religiosi, diplomatici, aziende	Spionaggio persistente, campagne globali APT, criminalità informatica	Topic 1, 3 e 5
5	2	Caso DNC e WADA – attacchi mirati russi per influenza politica e reputazionale	Attori statali e quasi-statali, spionaggio, sabotaggio, disinformazione	Topic 5 (e parte di 1/2/4)

Conclusioni

A dark blue background featuring a subtle, glowing blue digital wave pattern composed of small dots, creating a sense of depth and motion.

Dove bisogna migliorare

L'analisi effettuata evidenza che sebbene le misure di sicurezza informatica nelle imprese europee e italiane siano generalmente buone, ci sono ancora margini di miglioramento soprattutto in aree come la formazione dei dipendenti, la gestione del rischio umano e l'adozione di coperture assicurative contro gli incidenti informatici



La **digitalizzazione**, l'uso crescente di dispositivi **IoT**, il **cloud computing** e la **connettività costante** espongono le aziende a vulnerabilità crescenti. Inoltre le minacce informatiche diventano sempre più sofisticate, spesso legate a gruppi con finalità geopolitiche o di spionaggio industriale

Ruolo dell'intelligenza artificiale

L'intelligenza artificiale non può essere esclusa dal panorama della sicurezza informatica, in quanto rappresenta un alleato potente nella rilevazione di attacchi e nell'automazione dei controlli di sicurezza.

Tuttavia, se da un lato offre opportunità, dall'altro può essere utilizzata dagli attaccanti per orchestrare intrusioni sempre più mirate. Le IA, infatti, sono ormai in grado di generare codice malevolo, diventando una minaccia sempre più concreta, che rende le attività criminali accessibili a un numero crescente di attori.

È essenziale che l'integrazione dell'IA nella cybersecurity sia guidata da una solida governance etica, da competenze specialistiche e da rigorose politiche di controllo.



GPT-4 scrive un exploit prima del PoC pubblico. La corsa alla Patch non è mai stata così essenziale

Le imprese e in generale le istituzioni devono quindi:

- Investire nella formazione continua dei dipendenti;
- Integrare strumenti assicurativi nei piani di gestione del rischio;
- Rafforzare infrastrutture e procedure IT, con simulazioni e audit regolari;
- Collaborare a livello nazionale ed europeo, seguendo le linee guida UE per la cybersecurity.

La cybersecurity non è solo un costo, ma un investimento strategico fondamentale per proteggere il futuro delle imprese, sempre più dipendenti dai dati, e per prevenire situazioni che possano compromettere la sicurezza nazionale.

Inchiesta hacker, politici italiani con un trojan: le ipotesi degli inquirenti

Scoperto un vasto sistema di spionaggio in Italia con l'uso di un trojan RAT (un particolare tipo di malware) per accedere a dati riservati nel database del Viminale. Vittime del furto di dati sono varie figure della politica italiana, tra cui Ignazio La Russa, Matteo Renzi e Sergio Mattarella, di cui è stato hackerato l'indirizzo e-mail.



I Semafori Hackerati Parlano con la voce di Jeff Bezos ed Elon Musk. Tutta colpa di password banali

Bibliografia e sitografia

- Cybersecurity, cos'è e perché è importante - <https://www.agendadigitale.eu/tag/cyber-security/>
- Sasse, M.A., Brostoff, S., & Weirich, D. (2001). *Transforming the 'Weakest Link'—A Human/Computer Interaction Approach to Usable and Effective Security*. *BT Technology Journal*, 19(3), 122–131.
- ISO 31000:2018 - Gestione del rischio - Linee guida <https://www.cybersecurity360.it/legal/il-risk-management-e-la-nuova-iso-310002018-le-linee-guida/>
- GPT-4 scrive un exploit prima del PoC pubblico. La corsa alla Patch non è mai stata così essenziale, Redazione RHC - <https://www.redhotcyber.com/post/gpt-4-scrive-un-exploit-prima-del-poc-pubblico-la-corsa-all-patch-non-e-mai-stata-così-essenziale/>
- I Semafori Hackerati Parlano con la voce di Jeff Bezos ed Elon Musk. Tutta colpa di password banali. Redazione RHC - <https://www.redhotcyber.com/post/i-semafori-hackerati-parlano-con-la-voce-di-jeff-bezos-ed-elon-musk-tutta-colpa-di-password-banali/>
- Amaturo, E., Aragona, B., Grassia, M.G., Lauro, C.N., Marino, M. *Statistica per le scienze sociali*. UTET
- Kulshrestha, R., A Beginner's Guide to Latent Dirichlet Allocation(LDA), A statistical model for discovering the abstract topics aka topic modeling. - <https://medium.com/data-science/latent-dirichlet-allocation-lda-9d1cd064ffa2>
- Servidio, G., Inchiesta hacker, politici italiani con un trojan: le ipotesi degli inquirenti - <https://www.geopop.it/inchiesta-hacker-politici-italiani-con-un-trojan-le-ipotesi-degli-inquirenti/>

Your password is like your toothbrush

Don't share it and change it from time to time

Grazie per l'attenzione