# Progetto Machine Learning

Nicolò Fiorentini 933157 Anoir Kadmiri 933097

Università degli Studi di Milano-Bicocca

17 giugno 2025

# Dominio di riferimento

Il dominio di riferimento è la diagnosi di tumore al seno. Questo tipo di tumore è tra i più comuni tra le donne a livello mondiale.

# Obiettivi

**Obiettivo:** è classificare correttamente la natura del tumore al seno, distinguendo tra benigno e maligno.

# Features del dataset

- ~~id~~
- radius_mean
- texture_mean
- perimeter_mean
- area_mean
- smoothness_mean
- compactness_mean
- concavity_mean
- concave points_mean
- symmetry_mean
- fractal_dimension_mean
- radius_se
- texture_se
- perimeter_se
- area_se
- smoothness_se
- compactness_se
- concavity_se
- concave points_se
- symmetry_se
- fractal_dimension_se
- radius_worst
- texture_worst
- perimeter_worst
- area_worst
- smoothness_worst
- compactness_worst
- concavity_worst
- concave points_worst
- symmetry_worst
- fractal_dimension_worst

# Features del dataset

- ~~id~~
- radius_mean
- texture_mean
- perimeter_mean
- area_mean
- smoothness_mean
- compactness_mean
- concavity_mean
- concave points_mean
- symmetry_mean
- fractal_dimension_mean
- radius_se
- texture_se
- perimeter_se
- area_se
- smoothness_se
- compactness_se
- concavity_se
- concave points_se
- symmetry_se
- fractal_dimension_se
- radius_worst
- texture_worst
- perimeter_worst
- area_worst
- smoothness_worst
- compactness_worst
- concavity_worst
- concave points_worst
- symmetry_worst
- fractal_dimension_worst

# Target feature

| diagnosis value | Meaning |
|---|---|
| M | Malignant |
| B | Benign |

# Target feature

| diagnosis value | Meaning |
|---|---|
| 0 | Malignant |
| 1 | Benign |

# Analisi esplorativa

## Distribuzione del target

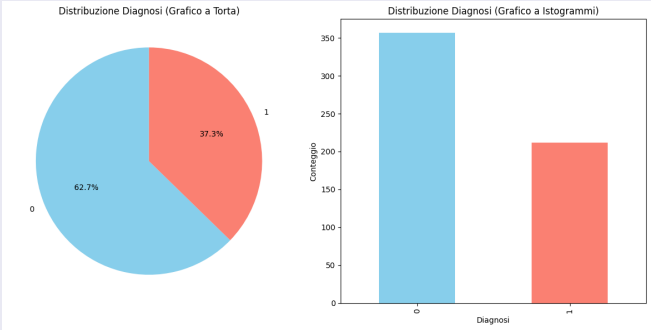


**Figure 1:** Distribuzione diagnosi

# Analisi esplorativa

## Distribuzione del target

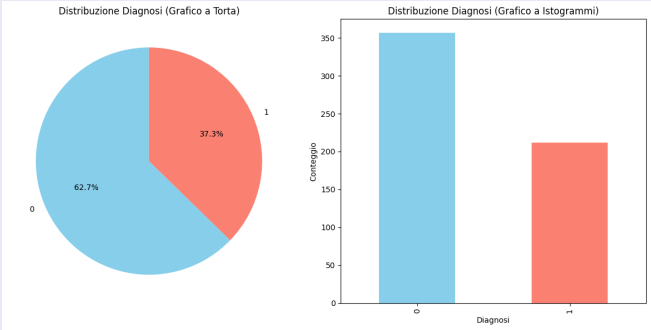


**Figure 2:** Distribuzione diagnosi

Abbastanza bilanciato

# Matrice di correlazione

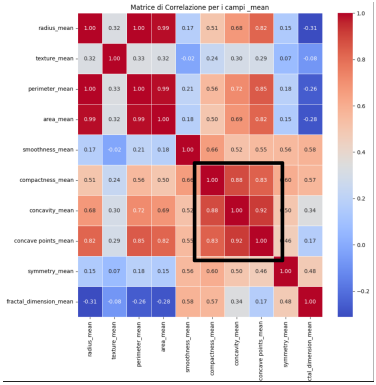


**Figure 3:** Matrice di correlazione per mean

L'area evidenziata con un riquadro nero mostra un gruppo di variabili altamente correlate

# Matrice di correlazione

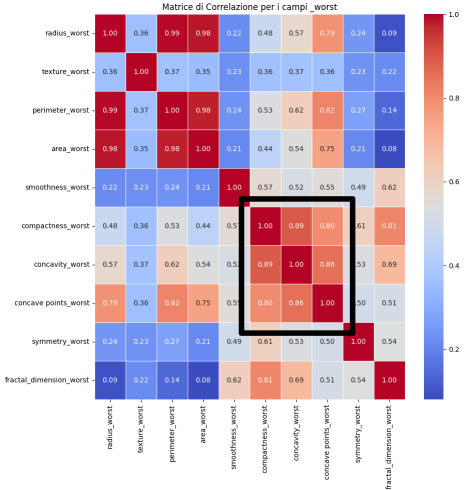


**Figure 4:** Matrice di correlazione per worst

# Matrice di correlazione

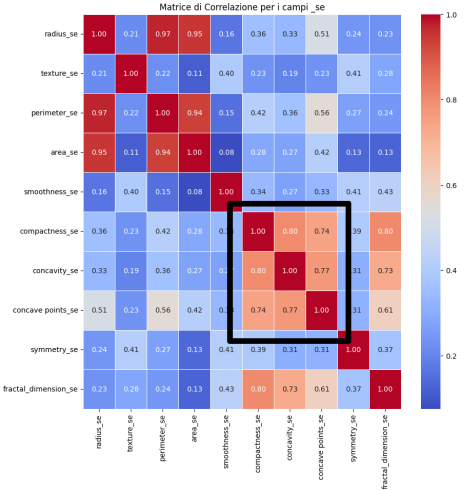


**Figure 5:** Matrice di correlazione per _se

# PCA

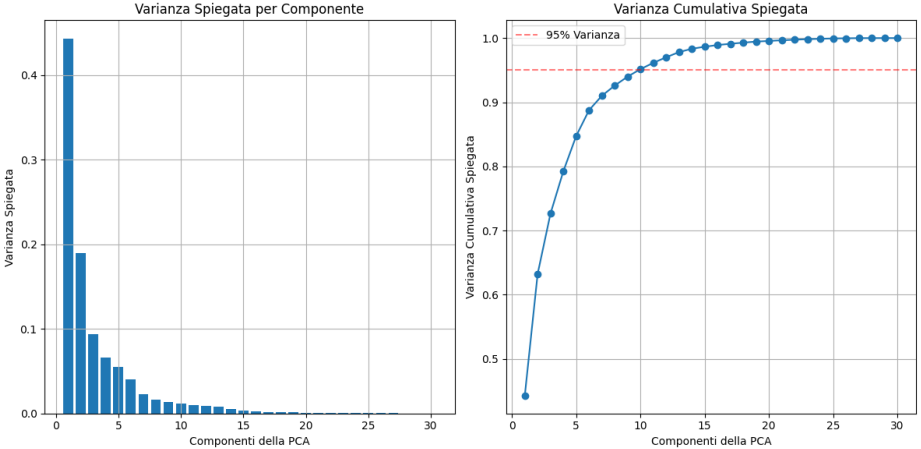


**Figure 6:** Grafico componenti

# PCA

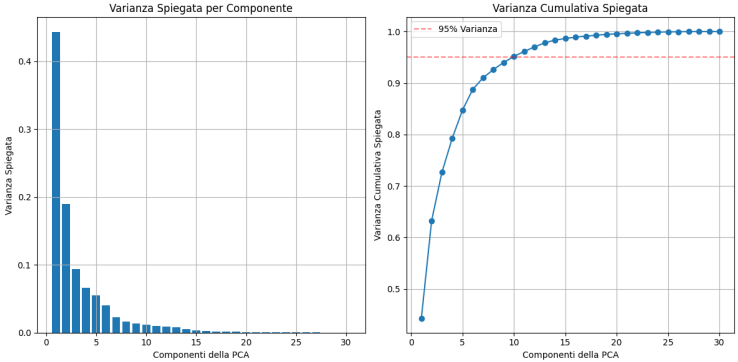


**Figure 7:** Grafico componenti

10 componenti per spiegare il 95% di varianza

# Reti neurali

2 reti neurali:

- reti neurali originale
- reti neurali con dati ridotti da PCA

# Reti neurali original

| Layer (type) | Output Shape | Param # |
|---|---|---|
| input_layer_14 (InputLayer) | (None, 30) | 0 |
| dense_28 (Dense) | (None, 20) | 620 |
| dense_29 (Dense) | (None, 1) | 21 |

**Figure 8:** Struttura finale rete neurale

# Reti neurali original

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Benigno | 0.98 | 0.99 | 0.99 | 249 |
| Maligno | 0.98 | 0.97 | 0.98 | 149 |
| accuracy | | | 0.98 | 398 |
| macro avg | 0.98 | 0.98 | 0.98 | 398 |
| weighted avg | 0.98 | 0.98 | 0.98 | 398 |

**Figure 9:** Score original reti neurali

# Reti neurali con features ridotti da PCA

| Layer (type) | Output Shape | Param # |
|---|---|---|
| input_layer_15 (InputLayer) | (None, 10) | 0 |
| dense_30 (Dense) | (None, 20) | 220 |
| dense_31 (Dense) | (None, 1) | 21 |

**Figure 10:** Struttura finale rete neurale

# Reti neurali PCA

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Benigno | 0.98 | 0.98 | 0.98 | 108 |
| Maligno | 0.97 | 0.97 | 0.97 | 63 |
| accuracy | | | 0.98 | 171 |
| macro avg | 0.97 | 0.97 | 0.97 | 171 |
| weighted avg | 0.98 | 0.98 | 0.98 | 171 |

**Figure 11:** Score PCA reti neurali

# Confronto reti neurali

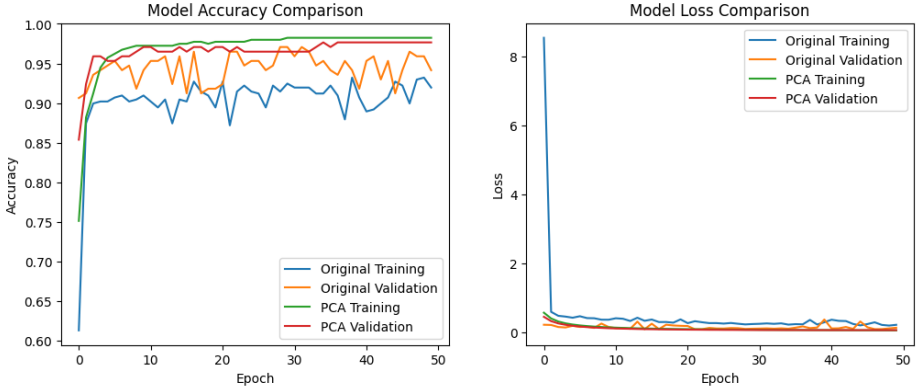


**Figure 12:** Confronto accuracy

# Confronto reti neurali

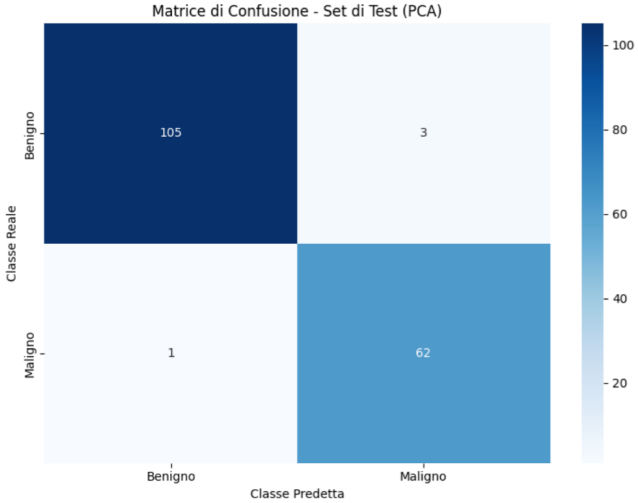


Figure 5.6: Matrice di confusione

# SVM

Due SVM:

- SVM originale;
- SVM ridotto con PCA;

# SVM

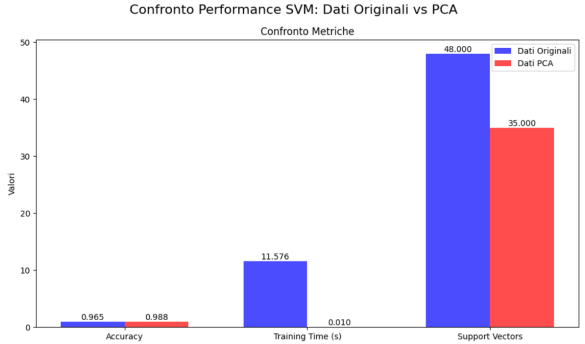


**Figure 14:** Confronto performance SVM

# SVM PCA

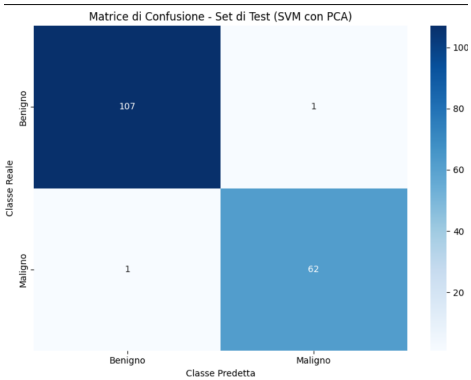


**Figure 15:** Confronto performance SVM

# SVM

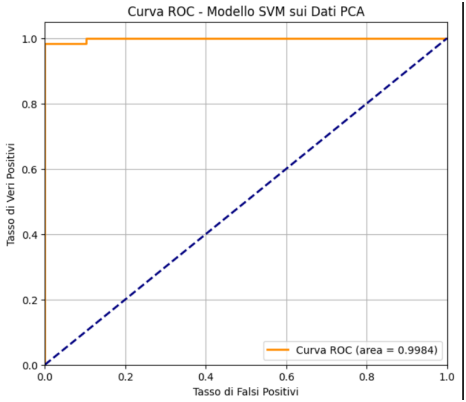


**Figure 16:** Curva ROC SVM

# CONFRONTO SVM E RETI NEURALI

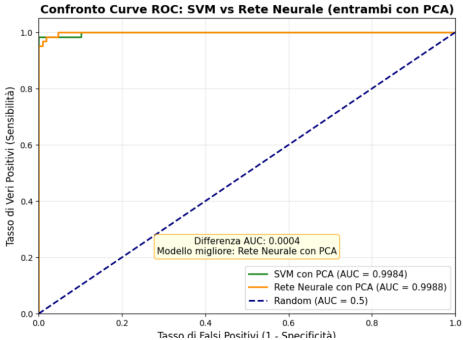


**Figure 17:** Curva ROC

# CONFRONTO SVM E RETI NEURALI

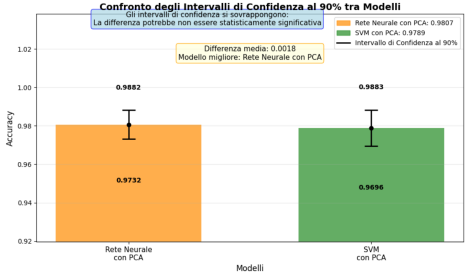


**Figure 18:** Intervallo confidenza

# CONFRONTO TEMPI

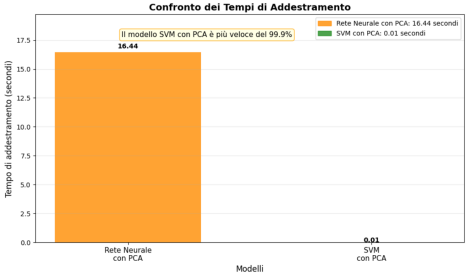


**Figure 19:** Confronto tempi