



**ITS** Academy  
meccatronico  
veneto

istituto tecnico superiore  
scuola speciale di tecnologia

---

## RELAZIONE 2

# Relazione 2

---

- La relazione va inviata a **battilana.matteo@gmail.com** entro il **04/04/2023**
- Giustificare e commentare ogni punto.
- Domande o parti **BONUS** sono opzionali: se è posto davanti alla domanda allora l'intera domanda è opzionale, se posto di seguito alla domanda allora solo quella parte della richiesta è opzionale;

# Relazione 2

---

- Formati accettati .docx, .odt, .pdf, .ipynb;
- Nome del file richiesto -> *CognomeNome\_Rel2.formato*
- Se si sceglie un formato **diverso da .ipynb** il codice può essere messo in allegato o messo a spezzoni all'interno della relazione;
- Se si sceglie il formato **.ipynb** assicurarsi che il codice sia già runnato e che quindi grafici ed altro siano già visibili: io non li runnerò ma li visualizzerò e basta.

# Relazione 2

---

La relazione si basa sul dataset -> [Breast Cancer](#)

## Features Analysis

1. Misurare l'eventuale sbilanciamento del dataset rispetto al target. Visualizzarlo con un grafico a torta o barre;
2. Calcolare la matrice di correlazione e ordinare le features secondo le più correlate al target;  
**BONUS ->** giustificare eventuali correlazioni forti fra le features e commentarle;
3. Valutare se aggiungere features aggiuntive date dall'interazione di features esistenti o dal loro sviluppo polinomiale;
4. Per le 10 features più correlate calcolare media, mediana, moda, deviazione standard e graficarne la distribuzione;  
**BONUS ->** graficare per ciascuna il box-plot;

# Relazione 2

---

## Bivariate Classification

5. Utilizzando le due features 'FractalDimensionWorst' e 'AreaMean' performare una classificazione binaria sul piano utilizzando i seguenti algoritmi: **DumbClassifier** (sempre classe più frequente), **LogisticRegression**, **LinearSVC**, **SVC** e **DecisionTreeClassifier**. Graficare le decision boundaries e valutare quale generalizza meglio in *accuracy* commentando;
6. Per il best performer, plottare la **precision-recall** curve. In questo caso cercheresti di ottenere una precision, recall o f1-score di almeno il 95%? Giustificare la strategia e riportare i valori delle metriche nel punto scelto.
7. Comparare gli algoritmi con la **ROC-curve** e determinare l'**AUC score**. Ordinare gli algoritmi secondo questa ultima metrica.

# Relazione 2

---

## Multivariate Classification

8. Utilizziamo ora tutte le features. Performare una classificazione binaria utilizzando i seguenti algoritmi: **DumbClassifier** (sempre classe più frequente), **LogisticRegression**, **LinearSVC**, **SVC** e **DecisionTreeClassifier**. Quale generalizza meglio in *accuracy*? Commenta;
9. Per il best performer, plottare la **precision-recall** curve. In questo caso cercheresti di ottenere una precision, recall o f1-score di almeno il 95%? Giustificare la strategia e riportare i valori delle metriche nel punto scelto.
10. Comparare gli algoritmi con la **ROC-curve** e determinare l'**AUC score**. Ordinare gli algoritmi secondo questa ultima metrica.

# Relazione 2

---

## Ensemble Methods

11. Utilizzando gli algoritmi precedenti nel caso multivariato, metterli assieme e valutare se **hard-voting** e **soft-voting** performano meglio dei singoli;
12. Allenare un modello **random forest** e valutarne le performances tarandosi sempre sul 95% della metrica che si è scelta;
13. **BONUS ->** Valutare quali features sono le più **importanti** per l'algoritmo precedente; graficarle in un grafico a barre ordinato per importanza decrescente.