Titolo

*Mancini Riccardo, Silvi Francesco, Tarsi Enrico.*

*Abstract*—Qualche riga su cosa conterrà la relazione.

# Introduzione

Il progetto svolto tratta un approfondimento nell’ambito probabilistico e dell’apprendimento automatico in Prolog. Questa è stato svoltoconsiderando lo stesso dataset di riferimento di pazienti blabla, che verrà di seguito approfondito….

La prima fase del progetto andrà a studiare l’aspetto probabilistico che permetterà di dare un'indicazione sulle relazioni fra i valori presenti nel dataset. Lo studio avverrà attraverso l'utilizzo del Cplint, con il quale verranno calcolate delle probabilità rilevanti (semplici, congiunte, condizionate e composte), queste consentiranno di studiare i legami fra i singoli attributi e la diagnosi del paziente. Questo studio andrà poi ad affiancare il risultato dell'apprendimento effettuato sul medesimo dataset, per valutarne la coerenza.  
L’apprendimento sviluppato servirà, oltre che per valutarne l’accuratezza di classificazione e di conseguenza il suo errore, ma anche per mettere a confronto due algoritmi (Shannon e …), due criteri (Entropia e Gini) di scelta dell’attributo nell’induzione dell’albero decisionale, rendendo allo stesso tempo gli algoritmi meno onerosi in termini di complessità computazionale.

…

# Dataset

L'Heart Failure Prediction dataset è una collezione di dati costituita da ben 746 casi di pazienti di cui si conoscono 11 caratteristiche (o features) che in un modo o nell'altro influenzano lo stato di salute del cuore di ognuno di essi. Di fatto, l'insufficienza cardiaca è un evento comune causato da malattie cardiovascolari, e grazie ai campioni raccolti in questo dataset è possibile prevedere una possibile malattia cardiaca. Infatti, le persone con malattie cardiovascolari o ad alto rischio cardiovascolare (per la presenza di uno o più fattori di rischio come ipertensione, diabete, iperlipidemia o malattie già accertate) necessitano di una diagnosi precoce e di una gestione tale da rendere un modello di machine learning di fondamentale importanza, ed è proprio per questo motivo che è stato ideato il dataset in questione.

Per quanto riguarda la struttura il dataset contiene, come già anticipato, 746 casi, ognuno dei quali descritto da ben 12 valori che possono assumere i vari attributi. Nello specifico gli 11 attributi, più l'ultimo attributo che definisce la classe (ossia la salute del paziente), sono:

1. **Age**: l'età del paziente (*anni*)
2. **Sex**: il sesso del paziente (*M*: maschio, *F*: femmina)
3. **ChestPainType**: tipo di dolore toracico (*TA*: angina tipica, *ATA*: angina atipica, *NAP*: senza dolore, *ASY*: angina asintomatica)
4. **RestingBP**: pressione sanguigna a riposo (*mm Hg*)
5. **Cholesterol**: colesterolo totale sierico (*mm/dl*)
6. **FastingBS**: glicemia a digiuno (*1*: se FastingBS > 120 mg/dl, *0*: altrimenti)
7. **RestingECG**: risultati dell'elettrocardiogramma a riposo (*Normal*: normale, *ST*: definisce un'anomalia dell'onda ST-T (inversione dell'onda T e/o elevazione o depressione dell'ST di > 0,05 mV), *LVH*: mostra un'ipertrofia ventricolare sinistra probabile o definita secondo i criteri di Estes)
8. **MaxHR**: frequenza cardiaca massima raggiunta (*valore numerico tra 60 e 202*)
9. **ExerciseAngina**: angina da sforzo (*Y*: si, *N*: no)
10. **Oldpeak**: sottoslivellamento ST, fa riferimento ad un'alterazione dell'elettrocardiogramma di superficie (*valore numerico tra -2.6 e 6.2*)
11. **ST\_Slope**: la pendenza del segmento ST di picco (*Up*: in salita, *Flat*: piatta, *Down*: in discesa)
12. **HeartDisease**: la classe di output (*1*: presenta problemi cardiaci, *0*: normale)

## Discretizzazione

Nel dataset alcuni attributi assumono valori continui o discreti ma comunque molto ampi per il nostro caso d'uso. Un esempio è l'età, che spazia dai 28 fino ai 77 anni, o il colesterolo che può potenzialmente assumere infiniti valori. Per questo è stato necessario intervenire con una discretizzazione del dataset, poiché le informazioni, derivanti da uno studio che tratta ogni singolo valore di un attributo come se stante, sarebbero potute essere forvianti.

Qui di seguito sono riportate le discretizzazioni applicate.

* **age**:
  + age ≤ 18 → *First*
  + 18 < age ≤ 60 → *Second*
  + age > 60 → *Third*
* **restingBP**:
  + restingBP < 120 → *Optimal*
  + 120 ≤ restingBP < 140 → *Normal/High*
  + 140 ≤ restingBP <160 → *High*
  + restingBP ≥ 160 → *Very high*
* **cholesterol**:
  + cholesterol < 200 → *Desiderable*
  + 200 ≤ cholesterol < 240 → *Moderately*
  + cholesterol ≥ 240 → *Extremely high*
* **maxHR**:
  + maxHR < 100 → *1*
  + 100 ≤ maxHR < 130 → *2*
  + 130 ≤ maxHR < 160 → *3*
  + maxHR ≥ 160 → *4*
* **oldpeak**:
  + oldpeak < 0.5 → *Low risk*
  + 0.5 ≤ oldpeak < 1.5 → *Normal risk*
  + oldpeak ≥ 1.5 → *High risk*

I restanti attributi non menzionati non avevano bisogno della discretizzazione, poiché lo erano di propria natura. In particolare sono:

* **sex**
* **chest\_pain\_type**
* **fastingBS**
* **restingECG**
* **excercise\_angina**
* **st\_slope**

Infine, per consentirne l'uso in Prolog è stata costruita la struttura degli esempi in modo tale che assumessero una forma esempio(Classe,[Attributi]), nel quale per "Classe" si intende lo stato di salute ("y" se malato, "n" se non malato), mentre per "[Attributi]" si intende una lista contenente tutti gli attributi con associato il relativo valore assunto nell'esempio.

Sono stati anche costruiti delle clausole per gli attributi nella forma attributo(NomeAttributo,[ValoriAttributo]) nel quale "[NomeAttributo]" identifica quale attributo si prende in esame, mentre "ValoriAttributo" è la lista contenente i valori che esso può assumere.

**(SOME EXAMPLES)**

# Analisi probabilistica (Cplint)

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Cplint e approfondimento delle librerie utilizzate

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Predicati probabilistici elementari

## Predicati probabilistici più complessi

## Rappresentazione grafica dei risultati

Qualche parola sulla libreria C3 e riportare qualche istogramma.

# Apprendimento automatico

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Tree Induction

## Criterio di scelta dell’attributo (Entropia vs gini)

## Risultati di classificazione

# Miglioramenti aggiuntivi

Qualcosa riguardo all’apprendimento totalmente sviluppato in Cplint… quello che diceva Dragoni, durante recivimento, di aggiungere alla presentazione durante il ricevimento.

##### Riferimenti

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.