

# Natural Language Processing nel campo Healthcare: storia della mia tesi

Luca Corbucci



# Chi sono?



## Luca Corbucci

- 🎓 Studente di Informatica presso l'Università di Pisa
- 🎙️ Podcaster
- 🦸 SuperHeroes Valley Community Manager



**Perché una tesi nel  
campo HealthCare?**

# Cos'è una “Electronic health record”

- Un EHR è un insieme di informazioni riguardanti la salute di pazienti
- Queste informazioni sono raccolte durante visite o ricoveri
- C'è uno standard per scrivere e organizzare EHR

# Vediamo un esempio

ROW_ID	SUBJECT_ID	HADM_ID	ADMIT TIME	DISCHTIME	ADMISSION_TYPE	ADMISSION_LOCATION	DISCHARGE_LOCATION	INSURANCE
21	22	165315	2196-04-09 12:26:00	2196-04-10 15:54:00	EMERGENCY	EMERGENCY ROOM ADMIT	DISC-TRAN CANCER/CHLDRN H	Private
22	23	152223	2153-09-03 07:15:00	2153-09-08 19:10:00	ELECTIVE	PHYS REFERRAL/NORMAL DELI	HOME HEALTH CARE	Medicare
23	23	124321	2157-10-18 19:34:00	2157-10-25 14:00:00	EMERGENCY	TRANSFER FROM HOSP/EXTRAM	HOME HEALTH CARE	Medicare
24	24	161859	2139-06-06 16:14:00	2139-06-09 12:48:00	EMERGENCY	TRANSFER FROM HOSP/EXTRAM	HOME	Private
25	25	129635	2160-11-02 02:06:00	2160-11-05 14:55:00	EMERGENCY	EMERGENCY ROOM ADMIT	HOME	Private

# Cosa abbiamo fatto?

**Doctor AI**



Previsioni su  
dati clinici di  
pazienti

**Doctor XAI**



Spiegazione  
delle previsioni

**Doctor XAI +  
Note cliniche**



Utilizzo delle note  
cliniche per  
arricchire le  
spiegazioni

# Come funziona Doctor AI

## INPUT:

Storia clinica del paziente

['562.12', '280.0', '211.3', '401.9',  
'250.00', '702.19', 'V10.3']

['562.12', '276.0', '250.00', '401.9',  
'V10.3']

['584.9', '276.5', '585', '532.90',  
'250.00', '285.9', 'V10.3', 'V44.2']

['569.69', '560.89', '998.59', '038.9',  
'995.91', '584.9', '585.9', '998.32',  
'250.00']



## OUTPUT:

Previsione delle  
patologie future:

158: Malignant neoplasm of  
retroperitoneum and peritoneum

259: Other endocrine disorders

238: Neoplasm of uncertain  
behavior of other and unspecified  
sites and tissues

# Come funziona Doctor XAI



Data l'istanza che  
vogliamo spiegare  
cerchiamo le più simili  
nel dataset



Vengono generate  
delle istanze simili e le  
facciamo classificare  
da Doctor AI



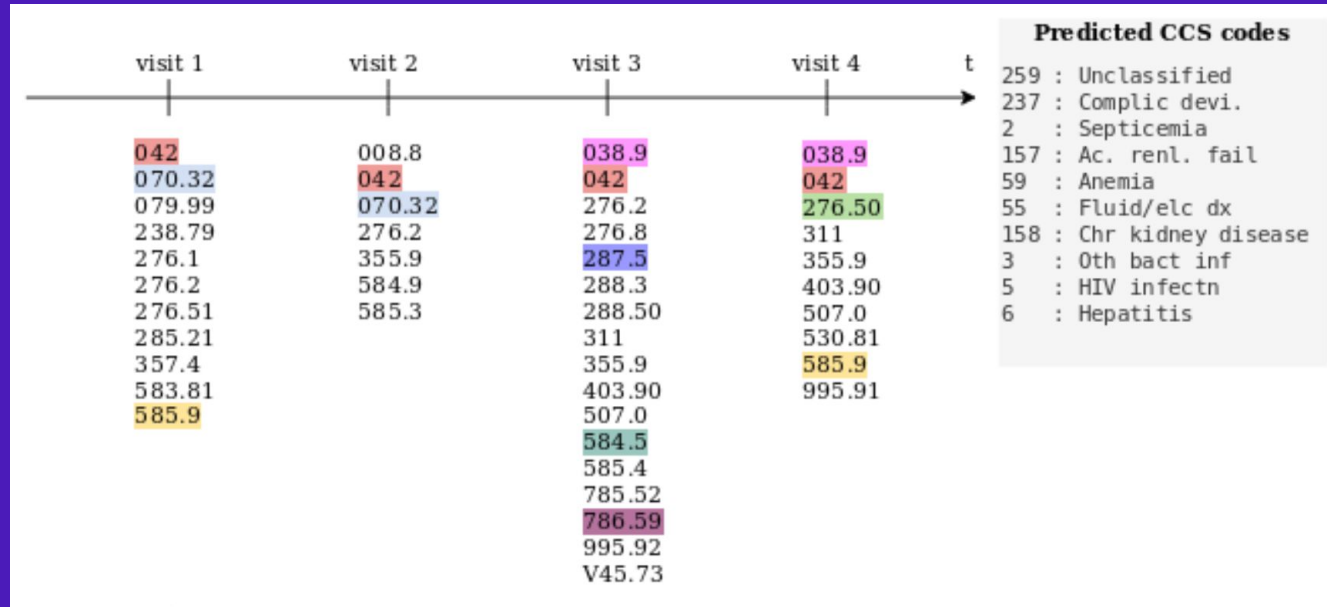
Viene creato un  
albero di decisione  
con i dati e con  
l'output di Doctor AI



Doctor XAI restituisce  
la regola che ha  
portato alla previsione  
di Doctor AI



# Come funziona Doctor XAI



# Sfruttiamo le note cliniche!

- I dataset medici contengono note cliniche scritte da dottori o infermieri
- Nelle note sono contenute informazioni sullo stato di salute del paziente
- Vogliamo estrarre dalle note cliniche le parti più rilevanti per la nostra spiegazione

# Come estraiamo le informazioni dalle note?



Utilizziamo l'ontologia medica  
SNOMED

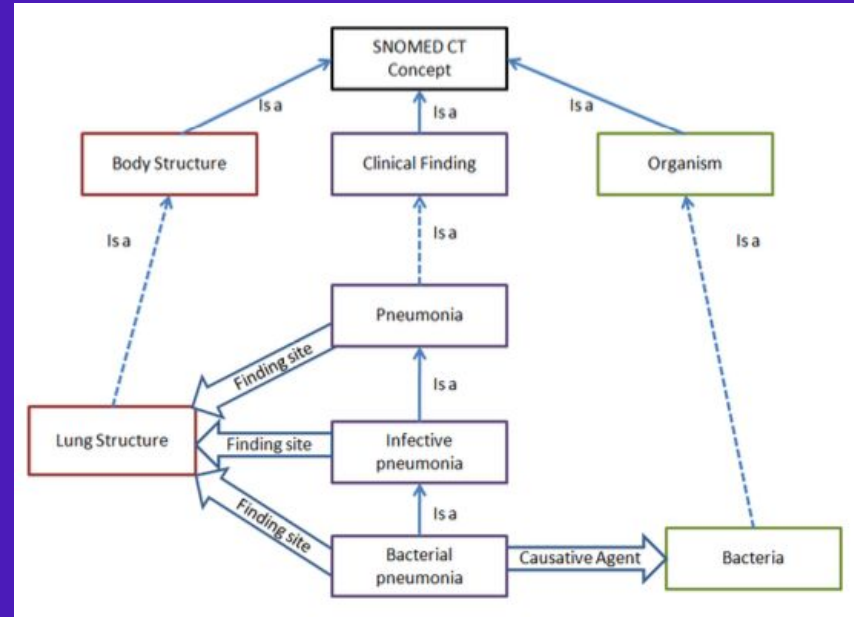


Per ogni nota evidenziamo le parti  
più simili alle relazioni prese  
dall'ontologia



L'obiettivo è di evidenziare per  
ogni malattia la descrizione, la  
causa, la parte del corpo e la  
morfologia associata

# SNOMED-CT



# In pratica?

Admission Date: [\*\*2163-9-21\*\*] Discharge Date: [\*\*2163-9-27\*\*]

Date of Birth: [\*\*2104-7-1\*\*] Sex: M

Service: MEDICINE

Allergies:  
Ceftriaxone

Attending:[\*\*First Name3 (LF) 943\*\*]

Chief Complaint:  
fever

Major Surgical or Invasive Procedure:  
multiple paracentesis

## History of Present Illness:

Pt is a 59 yo man w/ h/o Hep C cirrhosis s/p Liver xplant in [\*\*11-8\*\*], w/ chronic rejection (demonstrated on biopsy in [\*\*9-9\*\*]), recurrent Hep C on INF and ribavirin, B cell lymphoma, who p/w fevers, abdominal pain, SBP. Pt was in USOH until 1 week PTA when began feeling fatigued, had N/V approximately 1-2 episodes per day, non-bloody, non-bilious. 3 days PTA, pt began to have severe abdominal pain. He also noted increased abd girth, increased LE edema, R > L, denied any calf pain. Over past 3 days, pt also c/o cough with some sputum production, although difficult to bring up 2/2 abd pain. He also c/o laryngitis starting 3 days ago. ROS otherwise negative for BRBPR, melena, SOB, CP/pressure."

Descrizione: "Anemia in chronic kidney disease"



[0.314, -1.456, ....., 3.5644, 7.54432]

"Chronic Kidney Disease Stage II"



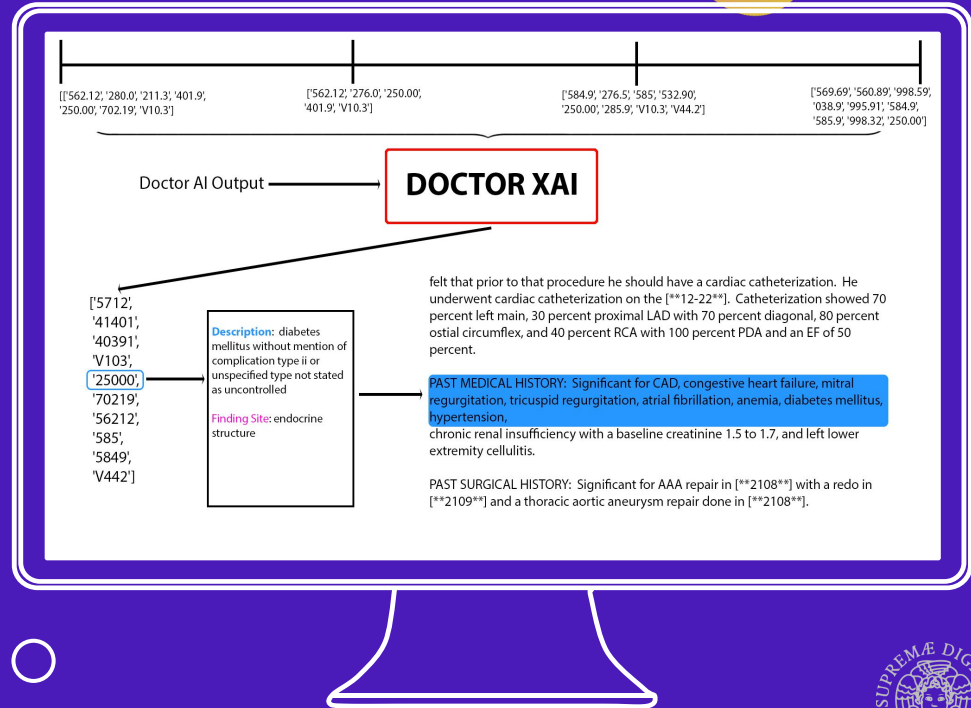
[0.436, 7.655, ....., -4.2533, 1.78824]

Calcolo la distanza tra le due rappresentazioni



# Obiettivo finale

Arricchire la spiegazione di Doctor XAI evidenziando le parti di nota più importanti.



# Grazie per l'attenzione!

