



Corso di Laurea Triennale in Informatica

# Identificazione di smell linguistici in Infrastructure as Code

**Prof. Fabio Palomba**  
**Prof. Dario Di Nucci**

**Danilo Aliberti**  
**Mat.: 0512103836**

✉ [d.aliberti10@studenti.unisa.it](mailto:d.aliberti10@studenti.unisa.it)

🐱 /woofz

in @danilo-aliberti



# Introduzione e Background

JADS Jheronimus  
Academy  
of Data Science



✉ d.aliberti10@studenti.unisa.it

🐾 /woofz

in @danilo-aliberti

Identificazione di smell linguistici in Infrastructure as Code

Danilo Aliberti

Università degli Studi di Salerno



## Obiettivi

Replicare uno studio precedente\* verificando le performance di Machine Learning per identificare smell linguistici in codice infrastrutturale

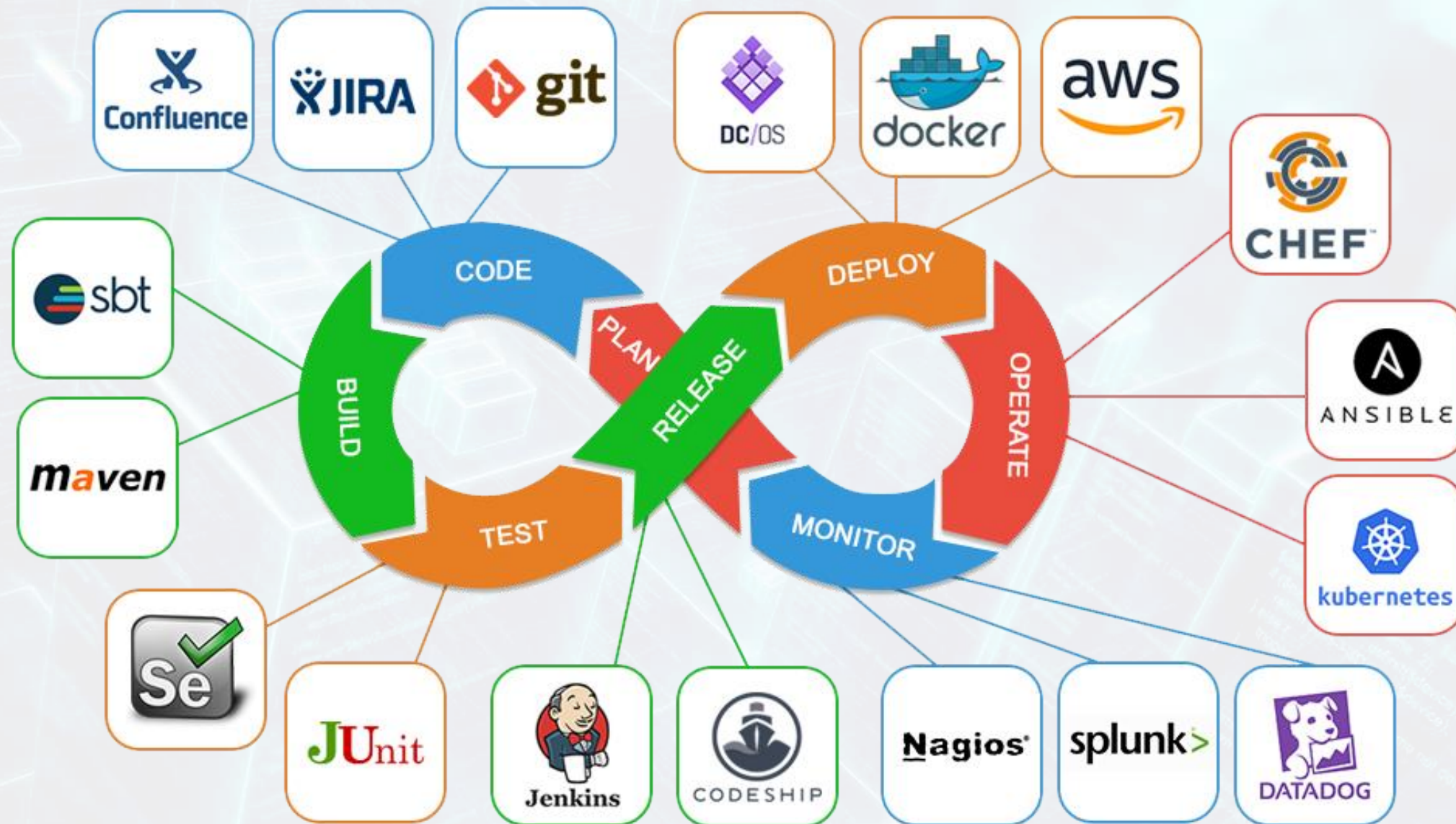


Esaminare le performance dei modelli su un nuovo dataset



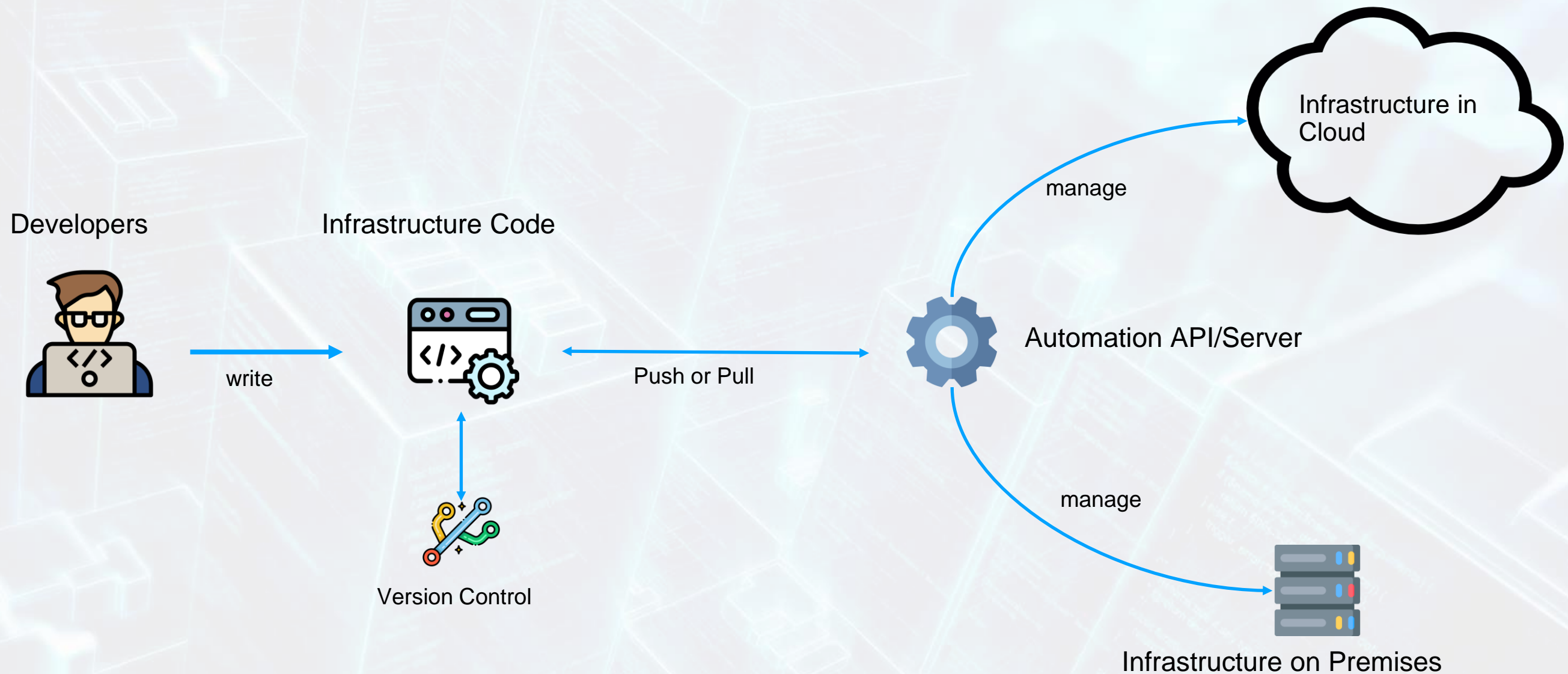
\* *FindICI: Using Machine-Learning to Detect Linguistic Inconsistencies between Code and Natural Language Descriptions in Infrastructure-as-Code. Empirical Software Engineering, to appear.*  
Borovits, N., Kumara, I., Di Nucci, D., Krishnan, P., Dalla Palma, S., Palomba, F., Tamburri, D. A., & van den Heuvel, W.-J. (2022)

## DevOps





## Infrastructure as Code: cos'è?



## Ansible Tasks

```
- name: Update web servers
  hosts: webservers
  remote_user: root

  tasks:
    - name: Ensure apache is at the latest version
      ansible.builtin.yum:
        name: httpd
        state: latest
    - name: Write the apache config file
      ansible.builtin.template:
        src: /srv/httpd.j2
        dest: /etc/httpd.conf

- name: Update db servers
  hosts: databases
  remote_user: root

  tasks:
    - name: Ensure postgresql is at the latest version
      ansible.builtin.yum:
        name: postgresql
        state: latest
    - name: Ensure that postgresql is started
      ansible.builtin.service:
        name: postgresql
        state: started
    - name: Ensure datadog-agent state stopped
      service:
        name: datadog-agent
        state: started
        enabled: true
```



ANSIBLE

## Inconsistenze linguistiche in IaC

```
- name: Update web servers
hosts: webservers
remote_user: root

tasks:
- name: Ensure apache is at the latest version
  ansible.builtin.yum:
    name: httpd
    state: latest
- name: Write the apache config file
  ansible.builtin.template:
    src: /srv/httpd.j2
    dest: /etc/httpd.conf

- name: Update db servers
hosts: databases
remote_user: root

tasks:
- name: Ensure postgresql is at the latest version
  ansible.builtin.yum:
    name: postgresql
    state: latest
- name: Ensure that postgresql is started
  ansible.builtin.service:
    name: postgresql
    state: started
- name: Ensure datadog-agent state stopped
  service:
    name: datadog-agent
    state: started
    enabled: true
```

Nome del Task — `- name: Ensure datadog-agent state stopped`

Corpo del Task — `service:`

`name: datadog-agent`

`state: started`

`enabled: true`

Il corpo del task **NON** rispetta il nome

Esempio di Ansible task non consistente





*\*nessun Ansible task è stato maltrattato nella stesura di questa tesi*

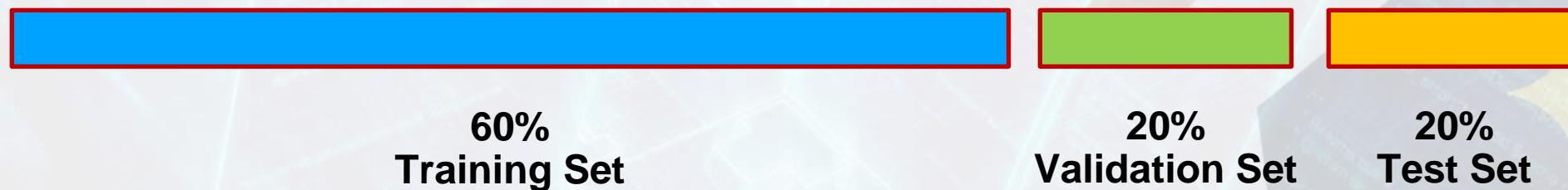


## Dataset

### Selezionato secondo criteri stabiliti

- Almeno l'11% dei files appartenenti alla repository devono essere script IaC;
- La repository ha almeno 10 contributors;
- La repository deve avere almeno due commit al mese;
- La repository non è un fork.

**10.396** Ansible Tasks



## Word Embedding

*fast*Text

Word2Vec

Doc2Vec

## Metriche

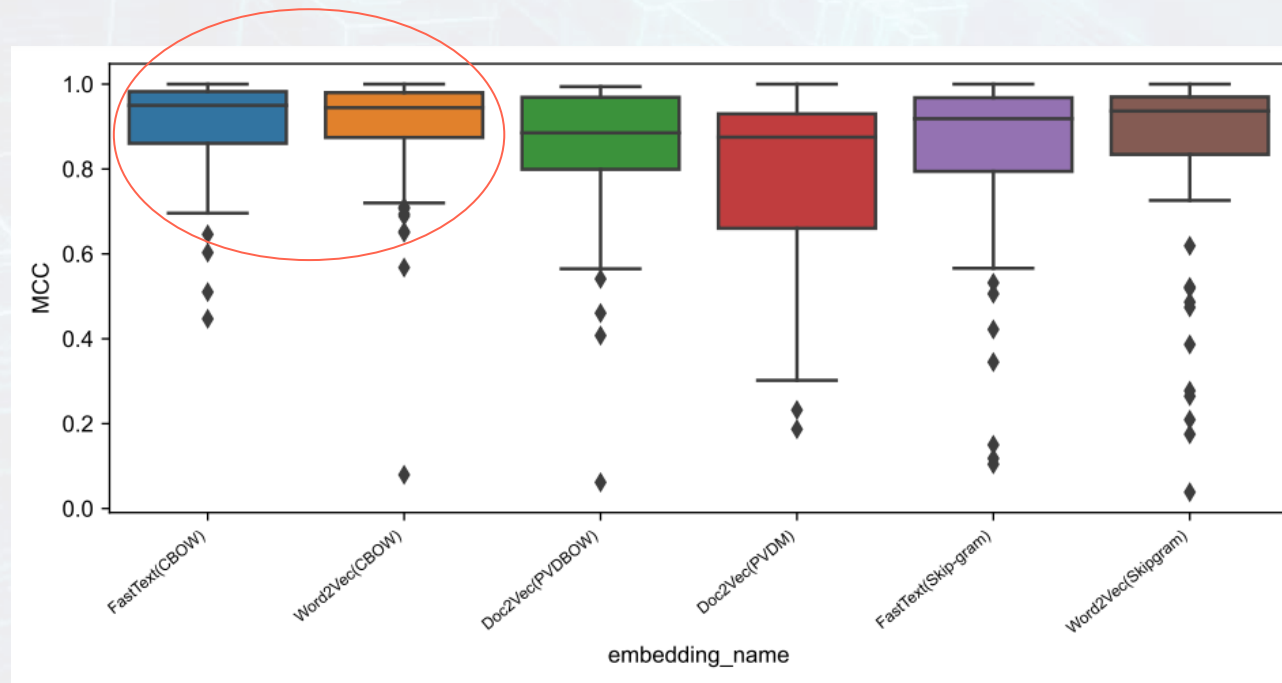




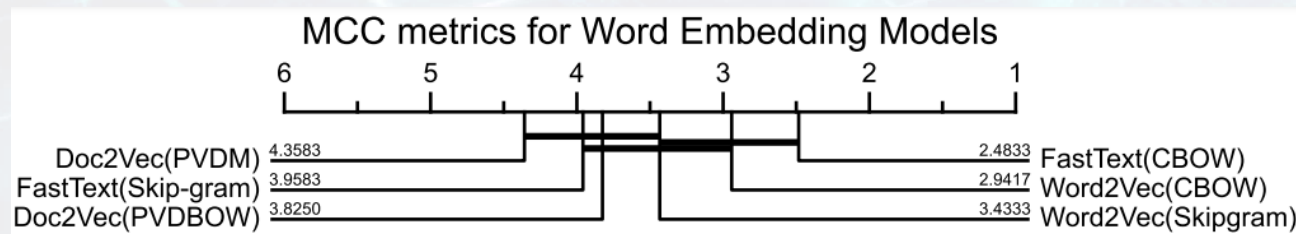
## Classificatori



# Qual è la migliore tecnica di Word Embedding?



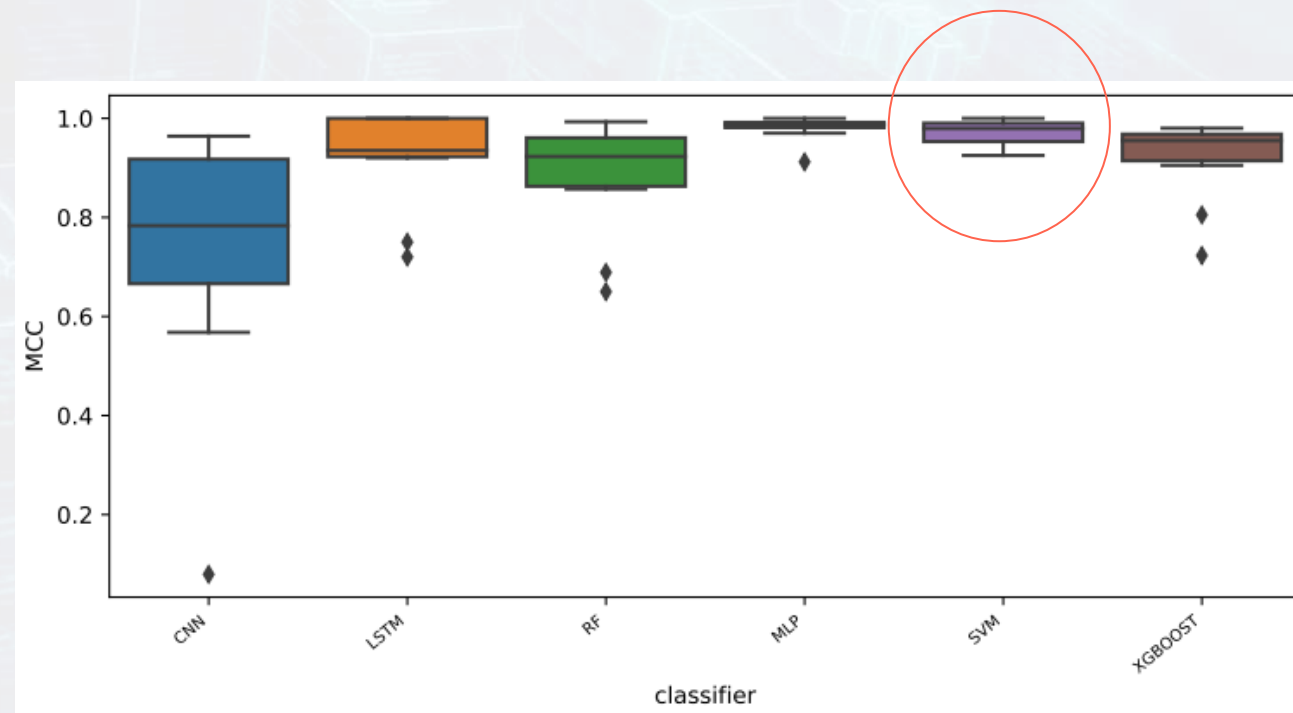
**FastText, Word2Vec**



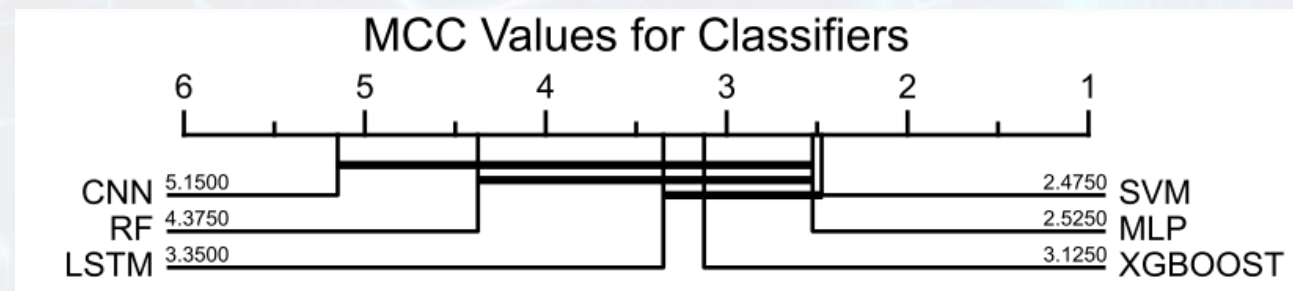
**Critical Difference diagram**



# Quale è il miglior classificatore?



**Support Vector Machine  
con Word2Vec**



**Critical Difference diagram**

## Introduzione e Background



### Obiettivi

Replicare uno studio precedente\* verificando le performance di Machine Learning per identificare smell linguistici in codice infrastrutturale



Esaminare le performance dei modelli su un nuovo dataset

\* FindCI: Using Machine-Learning to Detect Linguistic Inconsistencies between Code and Natural Language Descriptions in Infrastructure-as-Code. Empirical Software Engineering, to appear. Borovits, N., Kumara, I., Di Nucci, D., Krishnan, P., Dalla Palma, S., Palomba, F., Tamburri, D. A., & van den Heuvel, W.-J. (2022)

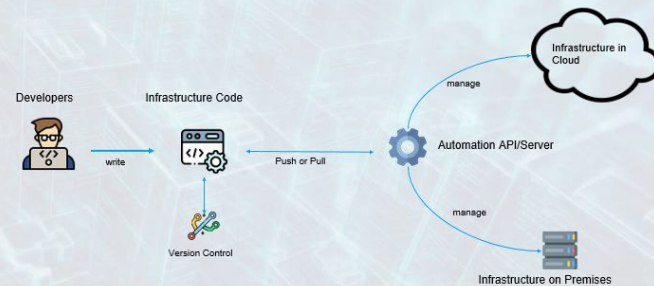
d.aliberti10@studenti.unisa.it  
woofz  
@danilo-aliberti

Identificazione di smell linguistici in Infrastructure as Code  
Danilo Aliberti  
Università degli Studi di Salerno

## Introduzione e Background



### Infrastructure as Code: cos'è?



d.aliberti10@studenti.unisa.it  
woofz  
@danilo-aliberti

Identificazione di smell linguistici in Infrastructure as Code  
Danilo Aliberti  
Università degli Studi di Salerno

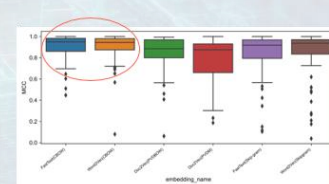
## Metodologia



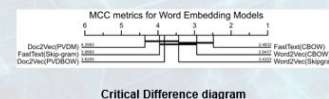
d.aliberti10@studenti.unisa.it  
woofz  
@danilo-aliberti

Identificazione di smell linguistici in Infrastructure as Code  
Danilo Aliberti  
Università degli Studi di Salerno

## Qual è la migliore tecnica di Word Embedding?



FastText, Word2Vec



d.aliberti10@studenti.unisa.it  
woofz  
@danilo-aliberti

Identificazione di smell linguistici in Infrastructure as Code  
Danilo Aliberti  
Università degli Studi di Salerno

# Identificazione di smell linguistici in Infrastructure as Code

Grazie per l'attenzione!



Questa tesi ha contribuito a piantare un albero in Ecuador!



*Il codice è interamente consultabile dalle repositories di riferimento su GitHub*

**Danilo Aliberti**

d.aliberti10@studenti.unisa.it

woofz

@danilo-aliberti