

# CODE SMILE: A ML-CSs's Detection Plugin

ESPERIMENTO CONTROLLATO

## INDICE

<b>OVERVIEW.....</b>	<b>2</b>
<b>PRE-SCREENING SURVEY.....</b>	<b>2</b>
METODO E SVILUPPO.....	3
<b>TESTING &amp; EVALUATION TOOL SURVEY.....</b>	<b>4</b>
CLASSIFICATION.....	4
METODO E SVILUPPO.....	4
TESTING & EVALUATION.....	8
METODO E SVILUPPO.....	8
<b>COMPARISON &amp; FEEDBACK SURVEY.....</b>	<b>9</b>
METODO E SVILUPPO.....	9

## OVERVIEW

In questa sezione è descritta com'è stato pensato il design dell'esperimento controllato per valutare come viene percepito e quanto è utile il tool dal punto di vista degli sviluppatori ML.

Si è deciso di procedere nei seguenti step, i quali verranno descritti e approfonditi nelle sezioni apposite:

1. **Pre-Screening Survey:** in questa sezione ai partecipanti viene chiesto di autovalutare la propria conoscenza ed esperienza in ambito Machine Learning e nello specifico in code smells in ML-Enabled Systems.
2. **Testing & Evaluation Tool Survey (Based on Phase):** in questa sezione, in base ai risultati del pre-screening, gli utenti vengono divisi in 2 gruppi: **Manuale-Tool** e **Tool-Manuale**  
Ogni gruppo effettuerà dei task di code smells detection su determinati file di esempio, prima usando il tool e poi no (e viceversa), compilando la sezione del survey dopo aver svolto i task (con tool o senza).
3. **Comparison Survey:** in questa sezione, l'utente dopo aver testato con tool (prima) e senza (poi) o viceversa, verrà chiamato a rispondere a delle domande sulla sua esperienza, andando ad indicare fattori come indice di gradimento con o senza, utilità del tool rispetto al non usarlo, precisione dello stesso rispetto ad un'analisi fatta senza tool ed altro...

Ogni sezione di questo documento è accompagnata dalla sotto-sezione "Metodo e Sviluppo" che descrive in base all'idea originale com'è stato sviluppato il prototipo di quella fase dello User Study.

## PRE-SCREENING SURVEY

In questa sezione, che rappresenta quella preliminare, i partecipanti allo studio sono chiamati a rispondere a domande sulle proprie conoscenze in ambito Machine Learning, con una particolare attenzione all'argomento "**Code smells**".

L'utente, oltre a descrivere il proprio background professionale, dovrà rispondere a domande sulle sue conoscenze nelle varie fasi di una pipeline ML e nello specifico sulla conoscenze di determinate librerie Python usate in progetti software ML

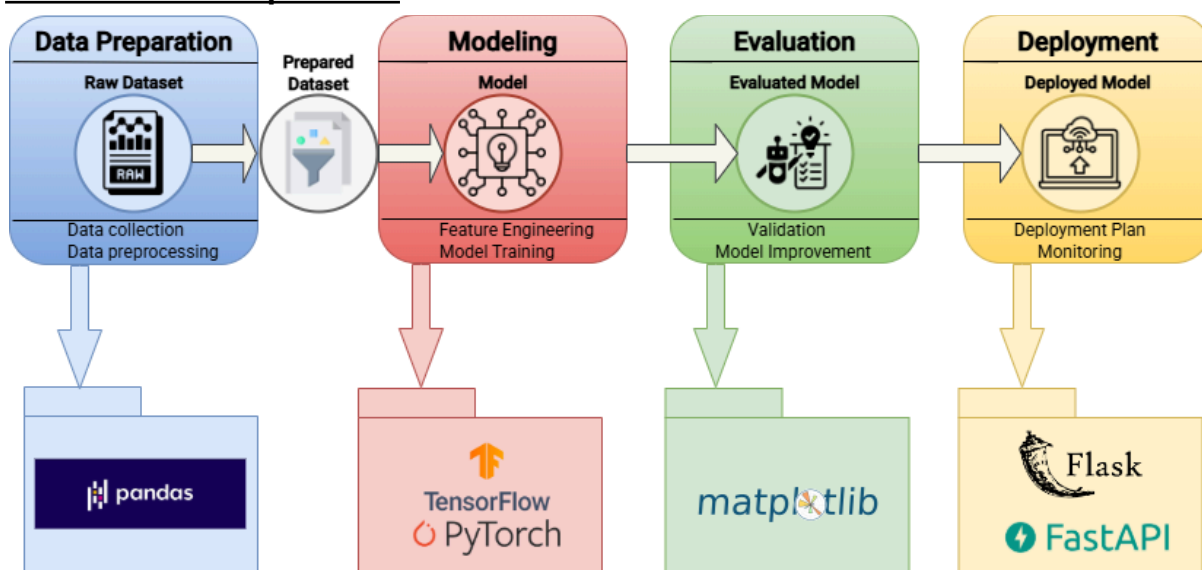
## METODO E SVILUPPO

Il link al form del Pre-Screening è il seguente:

[https://docs.google.com/forms/d/e/1FAIpQLSfw\\_FZL-X6FtYoVTNZxliJqgwG1LHCbZSlnAHv3VKrEzP-dLg/viewform?usp=dialog](https://docs.google.com/forms/d/e/1FAIpQLSfw_FZL-X6FtYoVTNZxliJqgwG1LHCbZSlnAHv3VKrEzP-dLg/viewform?usp=dialog)

Il form è diviso in 3 sezioni:

- **Anagrafica:** 2 domande inerenti al ruolo professionale occupato dal partecipante e da quanto tempo lavora in ambito Machine Learning (a risposta multipla)
- **Conoscenza ed Esperienza:**



Autovalutazione delle proprie conoscenze in ambito Machine Learning nelle fasi di una pipeline ML (come mostrato in figura sopra) (**Likert Scale**), autovalutazione della propria familiarità con librerie come Pandas, TensorFlow, Torch (**Likert Scale**), se si ha mai usato code smell detection tool (**T/F**) con opzionale domanda aperta se la risposta è "Sì" ad indicare quale nello specifico.

# TESTING & EVALUATION TOOL SURVEY

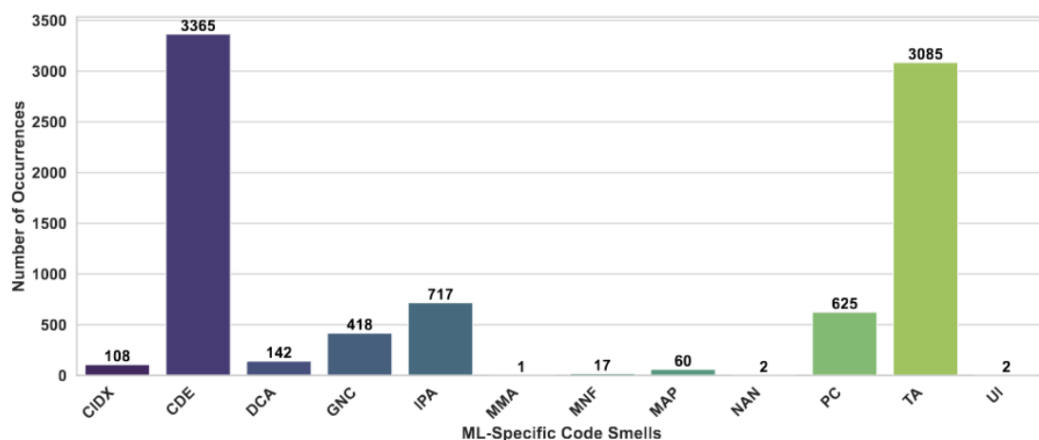
## CLASSIFICATION

Prima di procedere con il testing del Tool, vanno presi in considerazione i risultati del pre-screening e decidere in quale gruppo inserire il partecipante, ovvero nella Fase Manuale-Tool o quella Tool-Manuale e cosa fargli analizzare.

## METODO E SVILUPPO

26

Gilberto Recupito et al.



Basandosi sulla figura presente a pag. 26 del paper *“When Code Smells Meet ML: On the Lifecycle of ML-specific Code Smells in ML-enabled Systems”* dove vengono riportati i risultati di Code Smile in azione, si è deciso di scegliere i 6 più frequenti:

- CIDX
- CDE
- GNC
- PC
- TA
- IPA

Dopodiché è stato creato un file contenente uno smell specifico per ognuno dei smell indicati sopra.


File	Smell	Librerie coinvolte	Linee di codice d'apparenza dello smell
s1	CIDX	Pandas	10
s2	CDE	Pandas	4-7
s3	GNC	Torch	6
m1	PC	Torch	14
m2	IPA	Pandas	9
m3	TA	Tensorflow	6

Dopodiché sono stati creati 2 macro set di file che corrispondono alle 4 fasi dell'esperimento (T1,T2,M1,M2 dove T sta per Tool e M per Manual).

Questo perché si è deciso di seguire il seguente ordine/pattern di assegnazione dei partecipanti per fare la valutazione Tool-Manual dell'esperimento che viene illustrata dalla seguente tabella:

#### ASSEGNAZIONE PARTECIPANTI

Partecipante	Fase 1	Fase 2	File Fase 1	File Fase 2
P1	Tool 1	Manual 2	s1.py-s6.py	m1.py-m6.py
P2	Manual 1	Tool 2	m1.py-m6.py	s1.py-s6.py
P3	Tool 2	Manual 1	s1.py-s6.py	m1.py-m6.py
P4	Manual 2	Tool 1	m1.py-m6.py	s1.py-s6.py

 Da P5 ricomincia la rotazione d'accapo

I set sono così divisi:

### SET 1 (TOOL 1/MANUAL 1)

File	Smell	Rivelato?	Categoria	Note
s1.py	CIDX	✓	TP	Chain indexing diretto su DataFrame
s2.py	CDE	✓	TP	DataFrame con colonne ma senza dtype esplicito
s3.py	GNC	✓	TP	Manca zero_grad() prima di backward
s4.py	IPA (✗)	✓	FP	Segnalazione errata su drop(), uso corretto
s5.py	PC + TA	✗	FN	Uso di .forward() e tf.constant() in ciclo, non rilevati
s6.py	—	✗	TN	Distrattori: uso corretto di loc (CIDX-like), drop con inplace=True (IPA-like), zero_grad fuori ciclo (GNC-like)

### SET 2 (TOOL 2/MANUAL 2)

File	Smell	Rivelato?	Categoria	Note
m1.py	PC	✓	TP	Uso diretto di model.forward(), considerato scorretto
m2.py	IPA	✓	TP	Uso di drop() senza assegnazione o inplace
m3.py	TA	✓	TP	Concatenazione TensorFlow su tf.constant(), inefficiente
m4.py	CDE (✗)	✓	FP	Segnalato erroneamente come CDE, ma i dati non ambigui

m5.py	CIDX + GNC	✗	FN	Chain indexing separato su 2 linee + backward senza zero_grad
m6.py	—	✗	TN	Distrattori: uso di drop() con assegnazione (IPA-like), model(x) senza .forward() (PC-like), dtype specificato ma struttura ambigua (CDE-like senza violazione)

Come si evince dalle tabelle dei Set i primi 3 file sono dei **True Positive**, ovvero il Tool ha veramente rilevato lo smell corretto, il quarto file è un **False Positive**, il Tool ha individuato uno smell che, in realtà, non c'è.

Nei file 5 e 6 abbiamo che il Tool non rileva nulla, prima erroneamente, quindi si lascia scappare degli smells (**False Negative**) e poi correttamente (**True Negative**).

Questo permette di fornire insights interessanti per capire se l'utente riesce a notare le limitazioni comunque presenti nel Tool o meno, informazioni preziose per capire quanto più efficace può risultare lo strumento rispetto all'analisi manuale.

### Esempio di Mail per Assegnazione:

Ciao danyscp3602,

Sei stato assegnato alla seguente sequenza per il test **CodeSmile**:

 **Ordine delle fasi:** Tool → Manual

 **FASE MANUALE (M2):**

[Scarica i file manuali \(m1–m6\)](#)

 **FASE TOOL (T1):**

[Scarica i file con tool \(s1–s6\)](#)

 **Istruzioni:**

1. Scarica i file e svolgi l'analisi secondo l'ordine indicato
2. Identifica i code smells manualmente o con il tool
3. **Registra il tempo** per ciascun file

 Il tool è disponibile al link:

<https://github.com/Daniy2/CodeSmile>

 Una volta completato, compila il questionario:

<https://docs.google.com/forms/d/e/1FAIpQLSchRp-R1IBtY6poJUQxWIGKtIDzNeFY6Je5pkuFDQRbhgoHFQ/viewform?usp=dialog>

...

## NOTA TECNICA:

Si è resa possibile l'automatizzazione del processo di classificazione del partecipante basato sulle risposte del Pre-Screening Survey istantaneamente dopo la compilazione del form, grazie ad **Apps Script** (una piattaforma che consente di creare applicazioni per Google Workspace, la quale si basa su JavaScript e si integra con i servizi Google).

In questo caso è bastato collegare il file Google Sheets associato al Pre-Screening Survey (**Google Form**) con un uno script creato in Apps Script che si attiva, mediante un apposito trigger, ogni volta che l'utente compila ed invia il Google Form in questione.

## TESTING & EVALUATION

I due gruppi dovranno testare il tool, ma in che fase dell'esperimento cambia in base a quale gruppo si venga classificato.

Entrambi i gruppi sono chiamati a svolgere dei task di code smells detection.

I partecipanti di ogni gruppo avranno 2 file per fase su cui lavorare, verrà chiesto loro di risolvere il task apposito e di documentare il loro lavoro rispondendo a domande come: **numero smells trovati, tipo smells trovati, facilità e fiducia sul proprio task di rilevamento (solo per manuale), tempo impiegato, linee dove sono stati trovati gli smells, descrizione generale dell'esperienza...**

## METODO E SVILUPPO

Il survey di Evaluation è stato diviso in due sezioni:

- **FASE 1 (con tool** se l'ordine è T1-M2 oppure T2-M1, **manuale** l'ordine è M1-T2 oppure M2-T1), **FASE 2 (con tool** se l'ordine è M1-T2 oppure M2-T1, **manuale** se l'ordine è T1-M2 oppure T2-M1), dove viene chiesto quali smells sono stati trovati (**checkbox**) per ogni file, il tempo impiegato per detection per ogni file (**risposta aperta**), le linee riguardanti gli smell trovati (**risposta aperta**) commentare la propria esperienza (**risposta aperta**), il tempo impiegato (**risposta aperta**)

Per la fase con tool, c'è la parte di valutare l'esperienza d'uso con il tool e la domanda (**Likert Scale**) su quanto fosse d'accordo con il risultato dell'analisi e di eventualmente scrivere quali smell aveva individuato il partecipante a differenza del tool.

Per la fase senza tool, ci sono le domande su quanto il candidato si ritenesse sicuro del proprio lavoro (**Likert Scale**), e quanto fosse stato difficile trovare gli smells (**Likert Scale**).



- **FASE - VALUTAZIONE DEGLI SMELL E PERCEZIONE POST-ESPERIMENTO :**  
in questa fase, le domande sono volte a capire quale degli smell che il partecipante ha incontrato durante l'esperimento ha ritenuto più pericoloso o meno pericoloso, perché, a quale presterà maggiore attenzione e se aveva già avuto a che fare con questi smell.

Per visualizzare i survey, ecco i link dei form di compilazione:

- [T1-M2](#)
- [M1-T2](#)
- [T2-M1](#)
- [M2-T1](#)

## COMPARISON & FEEDBACK SURVEY

Arrivati alla parte finale, al candidato, indipendentemente dal gruppo in cui è stato classificato, verrà chiesto di compilare la sezione relativa alla comparazione delle due esperienze, ovvero gli verrà chiesto: **se si è trovato meglio con o senza tool e perché, che valutazione dà al tool, se lo userebbe come strumento nei suoi progetti, se lo ritiene utile, che voto dà a usabilità, efficacia, precisione, una domanda aperta su considerazioni e consigli per migliorare lo strumento.**

## METODO E SVILUPPO

La terza fase del survey si conclude con una sezione riguardante il confronto tra i due approcci e considerazioni finali.

È stato chiesto con la modalità descritta affianco a tra parentesi:

- Quale modalità gli fosse sembrata più efficace (**risposta multipla**)
- Con quale avesse rilevato più smells (**risposta multipla**)
- Se il tool lo avesse aiutato ad imparare qualcosa sull'argomento code smells (**risposta multipla**)
- Valutare singolarmente ognuno dei seguenti aspetti: usabilità, precisione, velocità e chiarezza dei messaggi (**Likert Scale**)
- Se consiglierebbe il tool a colleghi (**SI/NO**)
- Considerazioni finali e suggerimenti