

CODE SMILE: A ML-CSs's Detection Plugin

ESPERIMENTO CONTROLLATO

INDICE

| | |
|--|----------|
| OVERVIEW..... | 2 |
| PRE-SCREENING SURVEY..... | 2 |
| METODO E SVILUPPO..... | 3 |
| TESTING & EVALUATION TOOL SURVEY..... | 4 |
| CLASSIFICATION..... | 4 |
| METODO E SVILUPPO..... | 4 |
| TESTING & EVALUATION..... | 7 |
| METODO E SVILUPPO..... | 8 |
| COMPARISON & FEEDBACK SURVEY..... | 8 |
| 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

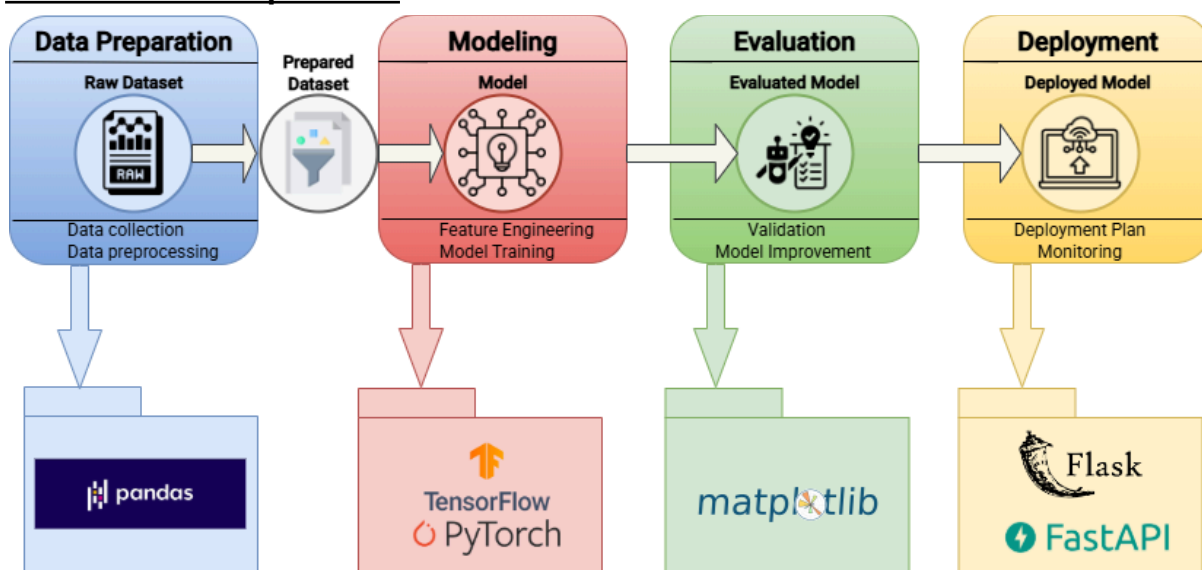
METODO E SVILUPPO

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

https://docs.google.com/forms/d/e/1FAIpQLSfw_FZL-X6FtYoVTNZxliJqgwG1LHCbZSlnAHv3VKrEzP-dLg/viewform?usp=dialog

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

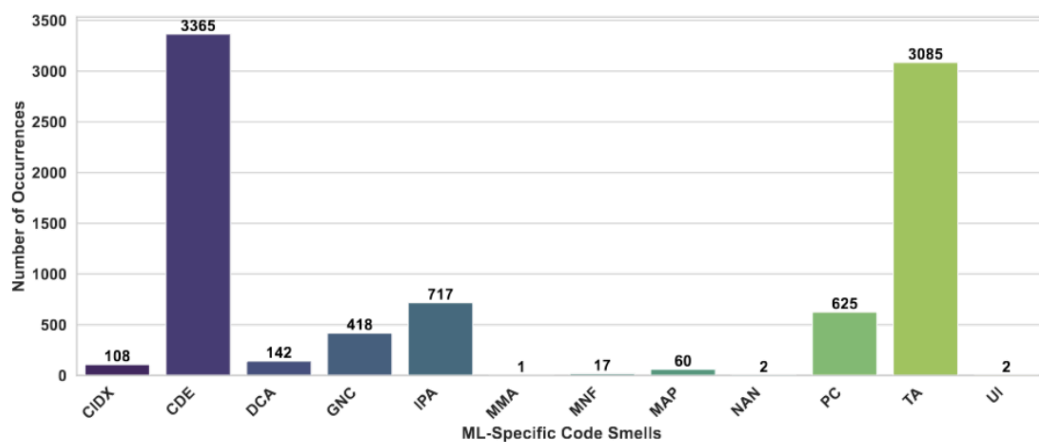
CLASSIFICATION

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 sono stati creati 6 file (in duplice copia e numerati dalla A alla F, ognuno in versione **_tool** e **_manual**) in questo modo:

| File | Smell | Librerie coinvolte | Linee di codice d'apparenza dello smell |
|-------|-------|--------------------|---|
| testA | CIDX | Pandas | 10 |
| testB | CDE | Pandas | 4-7 |
| testC | GNC | Torch | 6 |
| testD | PC | Torch | 14 |
| testE | IPA | Pandas | 9 |
| testF | TA | Tensorflow | 6 |

La scelta di averli in duplice copia permette che lo stesso file possa essere analizzato sia manualmente sia con tool (ovviamente da partecipanti diversi e in sessioni diverse), questo permette di avere insights significativi sull'utilità ed efficacia del tool perché possiamo vedere se effettivamente un determinato smell "scappa più facilmente" ad un utente o meno e ci permette di fare altre considerazioni.

In pratica, controlliamo questo fattore senza influenzare l'esperimento in nessun modo.

Dopo che l'utente ha compilato il survey del Pre-Screening gli vengono assegnati 2 file per la fase manuale e due per la fase con tool, seguendo questo schema di assegnazione:

Step 1 → Analizzare la familiarità con le librerie Python

- Vengono lette le risposte dell'utente sulla familiarità che ha sulle librerie Python (pandas, numpy, tensorflow, pytorch, scikit-learn)
- Vengono convertite queste risposte testuali in valori numerici (1-5)

Step 2 → Seleziona file potenzialmente candidati all'assegnazione

- manualCandidates → file con **almeno una** libreria conosciuta (≥ 3)

- toolCandidates → file con **nessuna** libreria conosciuta (<3)

Step 3 → Assegnazione file

Per quanto riguarda l'assegnazione, ci basiamo sul risultato dello step 2 e ci poniamo la seguente domanda:

Ci sono almeno 2 file manuali candidati per l'assegnazione e almeno 2 file tool candidati per l'assegnamento?

| Condizione | Scelta effettuata |
|-------------------|--|
| Risposta positiva | - Pesca 2 file random da manualCandidates - Pesca 2 file random da toolCandidates |
| Risposta negativa | - Pesca 2 file random dai candidati disponibili - Completa fase mancante usando file residui (senza duplicare quelli già assegnati) |

Step 4 → Ordinamento fasi

Per l'ordinamento dei gruppi a cui assegnare il partecipante si fa questo ragionamento:

- Si parte con l'ordine Manuale → Tool
- Poi si alterna: "Tool → Manuale", "Manuale → Tool", etc.


Questa scelta è stata fatta per avere un bilanciamento nel numero di partecipanti partecipano alle fasi, così avremo praticamente lo stesso campione di partecipanti di un gruppo e dell'altro gruppo, ottimo per avere un campione bilanciato.


Viene mandata un'email appena il partecipante compila il form che lo avvisa riguardo a quale gruppo è stato classificato, con i file in allegato da scaricare, le istruzioni su cosa fare, il link alla repo del tool e al form da compilare.

L'email che riceve l'utente è di questo formato:


Ciao

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


 **Ordine delle fasi:** Manuale → Tool

 **FASE MANUALE:**


- testC.py → https://drive.google.com/uc?export=download&id=1LlfwcqfnCBjrVba1EY_oMmghkgEkyC2Z
- testD.py → <https://drive.google.com/uc?export=download&id=1wQ7rNogrgRGNLjK4m0KZLw0m3x-hGerD>



 **FASE TOOL:**

- testA.py → https://drive.google.com/uc?export=download&id=1BPiBTbtDCsv-CoPr6CKiQbh8_gNugr9Z
- testF.py → <https://drive.google.com/uc?export=download&id=1Drzr4ofNfd1Nds23g2SGyKkqSJ6TipzC>

 **Cosa devi fare:**

1. Scarica i file indicati per ciascuna fase
2. Analizza i file manualmente e con il tool seguendo l'ordine indicato
3. Prendi nota dei code smells che riesci a identificare
4. **IMPORTANTE!! Munisciti di un cronometro per misurare il tempo che impieghi per risolvere ogni task sul file.**

 Il tool è disponibile al seguente link:
<https://github.com/Daniy2/CodeSmile>

 Una volta completato il test, clicca qui per compilare il questionario di valutazione:
 <https://tally.so/r/wQXvPk>

Grazie per il tuo contributo!
— Il team CodeSmile

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 (manuale** se Gruppo Manuale-Tool, **con tool** se Gruppo Tool-Manuale), **FASE 2 (manuale** se Gruppo Tool-Manuale, **con tool** se Gruppo Manuale-Tool), 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 domanda di valutare l'utilità del tool nel detectare smells.

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 dello smell:**
in questa fase, in base ai file che sono stati assegnati all'utente da valutare manualmente, c'è una sezione (per ogni file) dove viene rivelato che smell ci fosse al suo interno, e poi vengono poste 3 domande all'utente per valutare la percezione del pericolo dell'utente riguardante lo smell, la frequenza con cui incontra questo smell nei suoi progetti, se presterà maggiore attenzioni riguardo il seguente smell.

Tutte e 3 le domande sono di tipologia **Likert Scale** a 5 punti.

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

- [MANUALE-TOOL](#)
- [TOOL/MANUALE](#)

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