



# Classification of developeppers based on git activity

---

Moret Jérôme  
Curty Pierre-Alain  
Delessert Armand

# Sommaire

---

1. Description
2. Procédure
3. Application
4. Défauts

1.

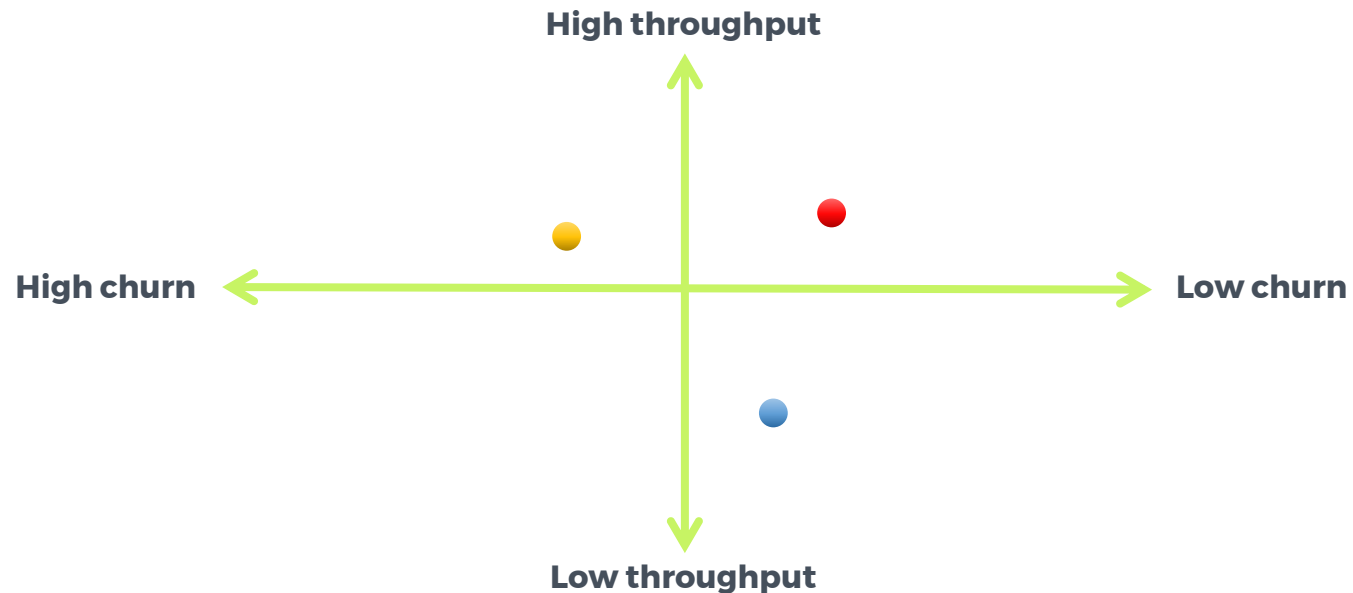
# Description



# Description

## ► Classification sur un espace à deux dimensions

- **Throughput** quantité de code ajouté
- **Churn** quantité de code réécrit



# 2.

## Procédure



# Procédure

Extraction des événements git

**git log --numstat**



Commit hash	Parent hash	Commit name	Author name	Commit date	Filename	Additions	Deletions	Churns
1	-	Write test.c	Jérôme Moret	18:00	Test.c	14	0	0
2	1	Modify test.c	Jérôme Moret	19:00	Test.c	3	3	?

churn ssi {  
Même auteur  
Dernière modif remonte à -3 semaines

# Procédure

Récupération des lignes affectées



Commit hash	Parent hash	Commit name	Author name	Commit date	Filename	Additions	Deletions	Churns
1	-	Write test.c	Jérôme Moret	18:00	Test.c	14	0	0
2	1	Modify test.c	Jérôme Moret	19:00	Test.c	3	3	?



**git diff --unified=0**

**1:test.c 2:test.c**



**[2,3,8]**

# Procédure

## Vérification des conditions de churn

Commit hash	Parent hash	Commit name	Author name	Commit date	Filename	Additions	Deletions	Churns
1	-	Write test.c	Jérôme Moret	18:00	Test.c	14	0	0
2	1	Modify test.c	Jérôme Moret	19:00	Test.c	3	3	3

[2,3,8]



**git blame -L2,+1**  
1 -- test.c



**Jérôme Moret** ✓  
**18:00** ✓



# Procédure Aggrégation

Commit hash	Parent hash	Commit name	Author name	Commit date	Filename	Additions	Deletions	Churns
1	-	Write test.c	Jérôme Moret	18:00	Test.c	14	0	0
2	1	Modify test.c	Jérôme Moret	19:00	Test.c	3	3	3
...	...	...	Delessert Armand	...	...	5	0	0
...	...	...	Delessert Armand	...	...	5	3	3
...	...	...	Curty Pierre-Alain	...	...	10	0	0
...	...	...	Curty Pierre-Alain	...	...	32	14	12



Author name	Additions	Churn	Churn rate
Jérôme Moret	17	3	0.18
Delessert Armand	10	3	0.3
Curty Pierre-Alain	42	12	0.28

 **Throughput**  
 **Churn**

3.

# Application

Méthodologie  
**Démonstration**  
Variantes

# Application Méthodologie



- ▶ Scripting Python
  - Librairie **GitPython**
  - Dataset.py
    - **Entrée** : chemin vers le dépôt
    - **Sortie** : dataset.csv
  - Aggregation.py
    - **Entrée** : dataset.csv
    - **Sortie** : result.csv
- ▶ Exploitation des résultats dans un tableur

# Application

Démonstration



- ▶ Dépôt GIT d'un projet bachelor
  - **Nom du projet** : EasyGoing
  - 5 contributeurs
  - 775 commits

# Classification of EasyGoing contributors

High throughput



● gweezer7

● michellesakam

● Raphaël racine

● Thibaud Duchoud

● edri

# Application Variantes



- ▶ Utilisation incrémentale
  - Dataset.py
    - s, --since <since\_date>
- ▶ Monitoring d'une équipe agile

# 4.

## Défauts



# Défauts

---

## ► Complexité

- Temps d'exécution de dataset.py sur **EasyGoing** : ~15min

## ► Solutions

- Réduire la complexité
  - Procédure
  - Algorithmes et structures de données
- Parallélisation
- Limite d'évènements



# Merci !

## Des questions ?

---

Moret Jérôme  
Curty Pierre-Alain  
Delessert Armand