

The Data Miners

TAED2

Samuel Cruz Gómez

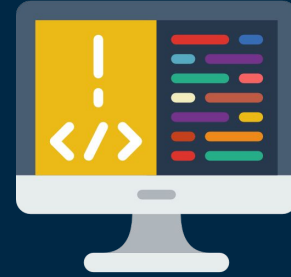
Àlec Lagarde Teixidó

Armand Alarcón Román

Business and Project goal

Code issues: Code smells,
anti-patterns

Help developers: Detecting
patterns

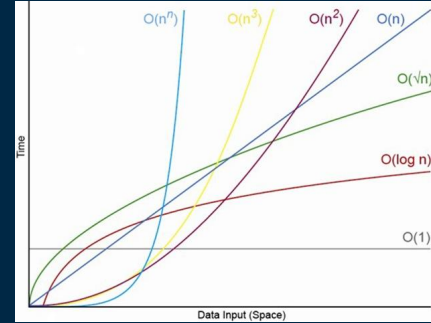


Success criteria

SonarQube issues: Complexity, commit hash

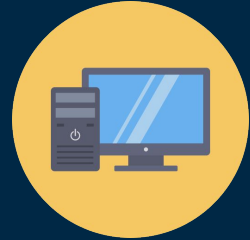
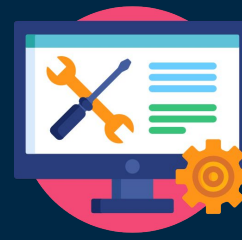
Model: Regression, Neural Networks

Results: Patterns, correlations



Assessment of the current situation

Inventory of resources:



Requirements:

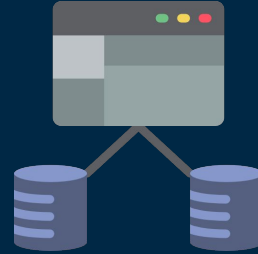
- 29th October
- Quality
- Reproducible

Assumptions:

- Resources
- Time

Assessment of the current situation

Constraints: - TechDebt dataset (v2.0)



Risks and contingencies:
- Difficulties finding patterns
- Redefine tasks or goals



Benefits from our analysis

- Efficiency
- Save resources



Project Plan



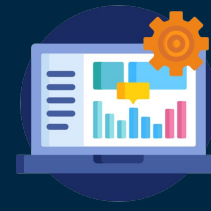
**Business
understanding**



**Data
understanding**



**Data
preparation**



Modeling



Evaluation



Deployment

Data understanding

Technical Debt Dataset

→ 10 .csv files

We will focus on the SonarQube-related tables, as our goal is to detect patterns in the SonarQube issues:

- SONAR_ANALYSIS
- SONAR_ISSUES
- SONAR_MEASURES
- SONAR_RULES

