SMA:
Software Modeling and Analysis

Practical Session

Week 11

Assignment 11

Discussion

A11 - Exercise 01 | Data in SDA

- ## What is a confusion matrix?

A confusion matrix is a table that shows the summary of a classification experiment outcome, i.e., the classification quality.

predicted class	actual class		
	cat	dog	rabbit
cat	A1 5	B1 2	C1 0
dog	A2 3	B2 3	C2 2
rabbit	A3 0	B3 1	C3 11

A11 - Exercise 01 | Data in SDA

- Which criteria are important when extracting the data?

Degree of automation.
The data extraction process should be automated as much as possible to avoid repetitive mistakes, e.g., a typo in the separation indicator in a CSV file could render data useless and might not be discovered immediately when only a subset of a large dataset is affected.

Availability of historical data.
It is recommended to keep any data, so that experiments can be rerun in case of implementation issues (mistake in the analysis) or for future verification (through different people).

…

A11 - Exercise 01 | Data in SDA

- Which are the three groups of software data?

Contextual data
The data related to the product or team (e.g., frequency of releases).

Constraint data
The data related to product constraints; the goals that need to be fulfilled (e.g., reliability of code).

Development data
The data related to the development process, the verification, and the code readiness (e.g., number of changed lines of code).

A11 - Exercise 02 | Process of SDA

- What is a use case for software data analytics?

There exist countless use cases. The most prevalent use case for SDA in industry is presumably the improvement of a product in terms of user experience, e.g., identifying inefficient code in order to apply some optimizations.

A11 - Exercise 02 | Process of SDA

- What are types of software analytics problems?

Bug measurements. Most bug measurements are incomplete since humans reported the occasional bugs manually. Thus, finding an indicator for bugs (e.g., the locality of bug prone classes) would greatly help to predict locations which need special attention.

Development practices. Developer practices might be suboptimal. For instance, one could improve the developer efficiency by analyzing how developers read comments and create guidelines for their optimal use.

Testing practices. It is impossible to consider every aspect of software testing. A way to go is to improve the existing testing practices. For instance, one could gather the amount of not sufficiently tested classes.

A11 - Exercise 02 | Process of SDA

- What is the recommended five-step guideline for working in software data analytics?

1. Problem identification
2. Collecting data
3. Descriptive analysis
4. Predictive analysis
5. Performance evaluation

You have to <u>attend the lecture</u> to reveal such slides.*

:-)

*Disclaimer:

The content that has been shown on this slide is irrelevant for the exam.

Assignment 12

Preview

A12 - Exercise 01 | Code smells

- What is the fundamental problem in developers bad code smell perception?

- What is "association rule mining" in the context of HIST?

- Provide a test code example that includes a test smell and explain.

A12 - Exercise 02 | Change of Roles

- Q01:   [Your question.]
  A01:   [Your corresponding answer.]

- Q02:   [Your question.]
  A02:   [Your corresponding answer.]

- Q03:   [Your question.]
  A03:   [Your corresponding answer.]

A12 - Exercise 03 | Exam Preparation

- Start preparing for the exam.

- Ask questions!