

# Apresentação da Evolução do Projeto de Formatura (1)

Aluno: Filipe Soares, 10774009

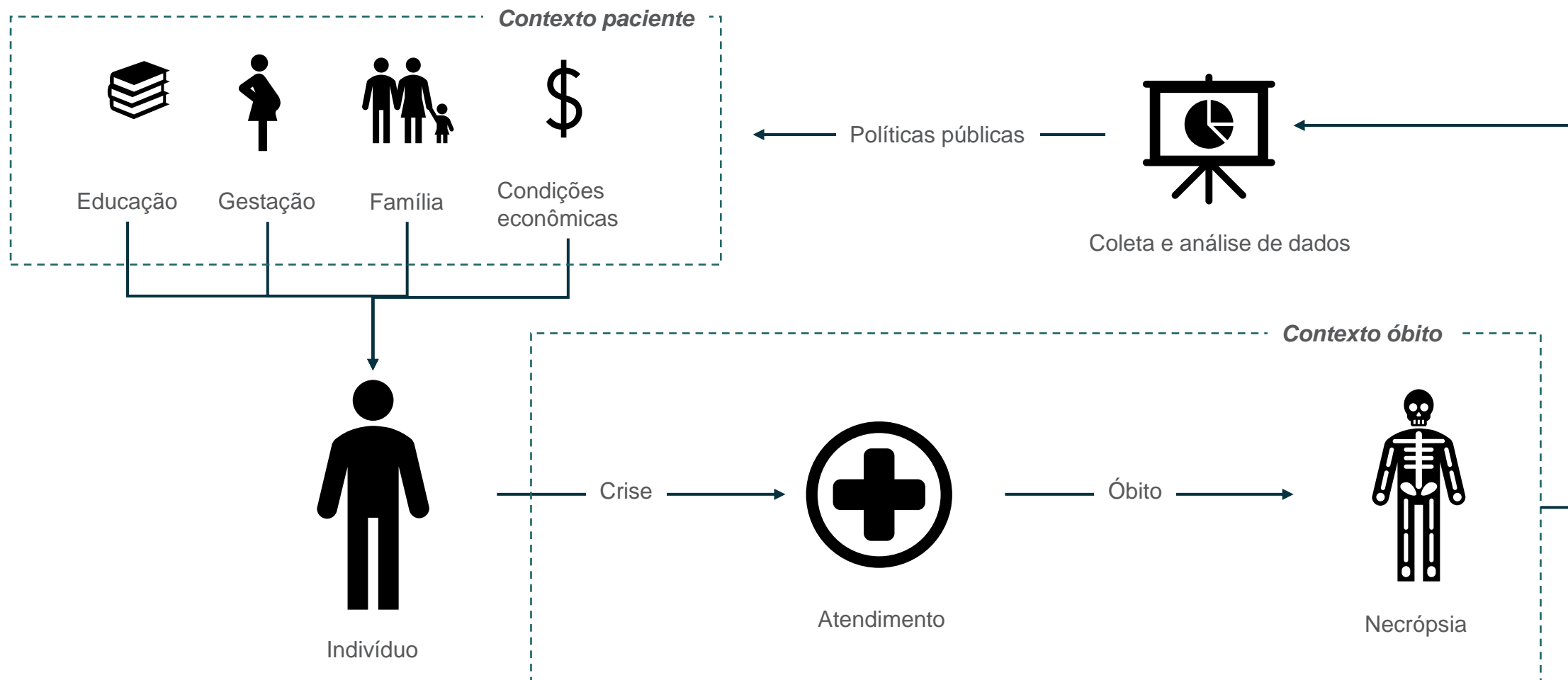
Orientador: Prof. Dr. Arthur Jordão

Engenharia de Computação

Poli-USP

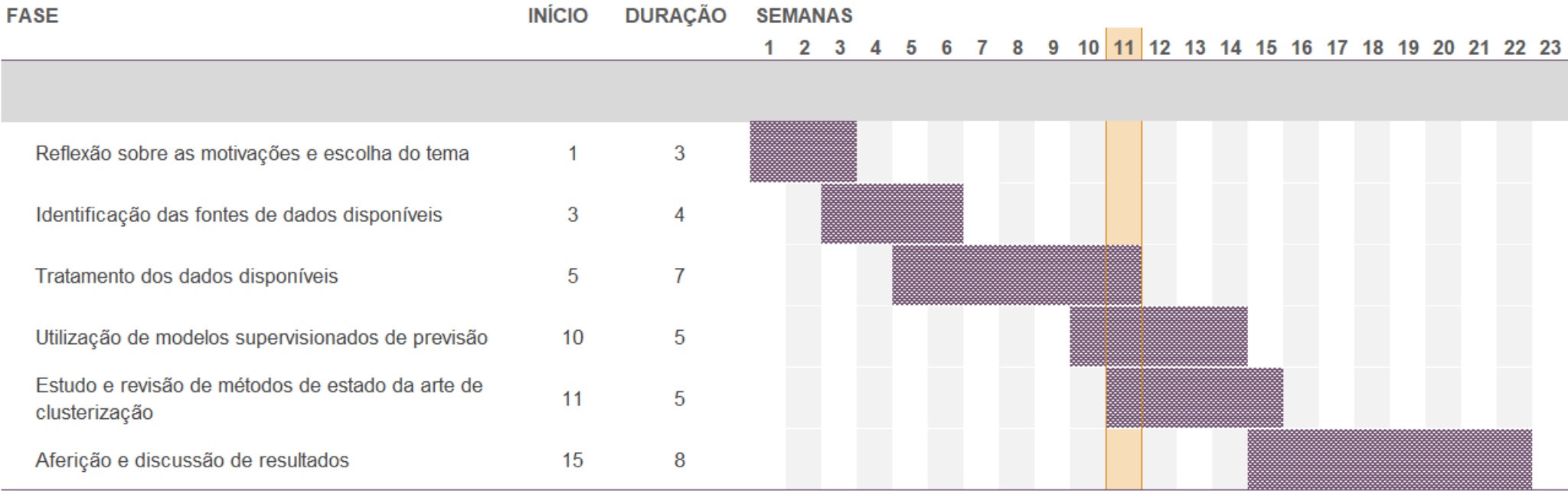
16 de outubro de 2023

O estudo da mortalidade é um recurso importante para a criação e orientação de políticas públicas



Na etapa atual do projeto, a base de dados já foi tratada e os modelo de supervisão e clusterização serão explorados nas próximas semanas

Fases



Atividades

- Estudar conceitos teóricos e práticos de aprendizado de máquina com ênfase no paradigma não-supervisionado
- Aplicar técnicas de clusterização nos dados disponíveis
- Conduzir experimentos e análise dos resultados
- Ler artigos científicos relacionados ao tema da pesquisa
- Organizar e documentar os códigos produzidos seguindo boas práticas de programação
- Redigir um documento técnico-científico (o trabalho final do projeto de formatura)

A principal fonte de dados para o projeto é a plataforma de ciência de dados aplicada à saúde da Fiocruz



SIM - Sistemas de Informação de Mortalidade

ETL (Extract, Transform and Load)



- 23.5 GB (702 arquivos) – 1 por ano por estado desde 1996
- Estado de SP: 5,52GB e 6.993.473 instâncias
- 164 parâmetros na base

Seleção de 41 parâmetros

- Filtro de parâmetros: (i) com quantidade de dados relevante, (ii) colunas redundantes e (iii) que fazem sentido para análise
- Codificação dos parâmetros

- Classificação dos parâmetros em categorias do paciente e do óbito