



Universidade do Minho

# Projeto 1

“Screen Time vs Mental Wellness Survey 2025”

TP2

Grupo 2

Aprendizagem Automática em Sistemas Empresariais

Mestrado em Engenharia e Gestão de Sistemas de Informação

Gonçalo Faria Sousa e Silva PG61458

Lara Sofia Freitas Esteves PG61470

Lucas Serralha dos Santos Peres PG60114

Simão Ribeiro da Silva PG60133

Docentes: Manuel Filipe Santos e Pedro Pereira

2025/2026

Guimarães, 2025



# Índice

<b>Índice .....</b>	<b>3</b>
<b>1. <i>Introdução</i>.....</b>	<b>4</b>
<b>2. <i>Business Understanding</i> .....</b>	<b>5</b>
<b>    2.1. Determinação dos Objetivos de Negócio.....</b>	<b>5</b>
2.1.1. Contexto de Negócio .....	5
2.1.2. Objetivos de Negócio .....	5
2.1.3. Critérios de Sucesso do Negócio .....	6
<b>    2.2. Avaliação da Situação .....</b>	<b>6</b>
2.2.1. Inventário de Recursos.....	6
2.2.2. Requisitos, Premissas e Restrições .....	8
2.2.3. Riscos e Contingências .....	9
2.2.4. Terminologia .....	10
2.2.5. Custos e Benefícios .....	11
<b>    2.3. Determinação das Metas de Mineração de Dados .....</b>	<b>12</b>
2.3.1. Metas de Mineração de Dados .....	13
2.3.2. Critérios de Sucesso Técnico .....	14
<b>    2.4. Produção do Plano do Projeto .....</b>	<b>15</b>
2.4.1. Plano do Projeto .....	15
2.4.2. Avaliação Inicial de Ferramentas e Técnicas .....	18



## 1. Introdução

No âmbito da unidade curricular de Aprendizagem Automática em Sistemas Empresariais, foi proposto o desenvolvimento de um projeto com o objetivo de estudar e analisar a relação entre o tempo de ecrã (screen time) e o bem-estar mental dos indivíduos, de forma a compreender como os hábitos de utilização de dispositivos digitais afetam o sono, o stress e a saúde mental. O projeto baseia-se no conjunto de dados “Screen Time vs Mental Wellness Survey – 2025”, disponibilizado na plataforma Kaggle.

O grupo segue a metodologia CRISP-DM (Cross Industry Standard Process for Data Mining), que fornece um modelo de orientação para projetos de *Data Mining* através de uma sequência de seis fases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation e Deployment. No entanto, neste Projeto serão desenvolvidas apenas as cinco primeiras fases.



## 2. Business Understanding

Nesta secção serão analisados os aspetos essenciais do projeto do ponto de vista do negócio. Serão abordados a Determinação dos Objetivos de Negócio, a Avaliação da Situação incluindo recursos, restrições, riscos e custos, a Determinação das Metas de Mineração de Dados traduzindo os objetivos de negócio em metas técnicas e a Produção do Plano do Projeto com definição das fases, duração, ferramentas e técnicas a utilizar.

### 2.1. Determinação dos Objetivos de Negócio

Nesta subsecção, procura-se compreender o que a clínica ou stakeholder pretende alcançar do ponto de vista do negócio, estabelecendo o contexto em que o projeto se insere, os objetivos principais e secundários, e os critérios que permitirão avaliar o sucesso do projeto.

#### 2.1.1. Contexto de Negócio

A Clínica de Cuidados Mentais procura compreender de que forma o uso excessivo de tecnologia e o tempo de ecrã diário influenciam o bem-estar psicológico dos seus pacientes. Com o aumento significativo da utilização de dispositivos digitais, têm surgido preocupações quanto ao impacto deste comportamento em aspetos como o nível de stress, a qualidade do sono e o equilíbrio emocional.

Neste contexto, a clínica pretende desenvolver uma solução analítica baseada em *Data Mining* e *Machine Learning* que permita identificar padrões de comportamento digital e prever potenciais riscos para a saúde mental. Esta análise permitirá à clínica oferecer serviços personalizados e estratégias de prevenção adaptadas ao perfil e aos hábitos de cada paciente.

#### 2.1.2. Objetivos de Negócio

O principal objetivo de negócio consiste em analisar e prever a relação entre o tempo de ecrã e o bem-estar mental, determinando os fatores que mais influenciam a qualidade do sono e o nível de stress. Pretende-se utilizar modelos preditivos para identificar indivíduos com risco elevado de desequilíbrio digital, permitindo à clínica adotar medidas de intervenção mais eficazes.

De forma complementar, o projeto visa apoiar a clínica na tomada de decisões baseadas em dados, contribuindo para uma melhor compreensão do comportamento digital dos pacientes e para o desenvolvimento de programas de acompanhamento personalizados.

### 2.1.3. Critérios de Sucesso do Negócio

O sucesso do projeto será avaliado pela capacidade de gerar informação relevante e açãoável para apoiar decisões clínicas. Os critérios de sucesso incluem:

- Obter um nível de precisão satisfatório nas previsões da qualidade do sono e do nível de stress;
- Identificar padrões claros que relacionem o tempo de ecrã com o bem-estar mental;
- Gerar insights comprehensíveis e úteis para os profissionais de saúde mental;
- Permitir à clínica reduzir casos de stress elevado ou má qualidade do sono através de estratégias personalizadas baseadas nos resultados do modelo.

## 2.2. Avaliação da Situação

Esta secção aprofunda a investigação sobre recursos, restrições e outros fatores que influenciarão o plano do projeto.

### 2.2.1. Inventário de Recursos

Esta secção descreve os principais recursos disponíveis para o desenvolvimento do projeto, incluindo dados, recursos computacionais e software a utilizar.

- **Dados:**

O principal recurso de dados é o dataset “Screen Time vs Mental Wellness Survey – 2025”, obtido na plataforma Kaggle.

O conjunto de dados é composto por 400 registos e 15 variáveis úteis. Cada linha representa um indivíduo, contendo informações acerca dos seus hábitos digitais, padrões de sono, níveis de stress, produtividade e bem-estar mental.



As variáveis incluem atributos demográficos (idade, género, ocupação), variáveis comportamentais (tempo total de ecrã, no trabalho e em lazer) e indicadores de saúde mental (qualidade de sono, stress, índice de bem-estar).

- **Recursos de Computação:**

O projeto será desenvolvido num ambiente computacional académico, recorrendo a:

- Computadores pessoais dos membros do grupo.
- Serviços de nuvem gratuitos.

Estes recursos são necessários para o desenvolvimento do projeto, não sendo necessária uma infraestrutura de alto desempenho.

- **Software:**

Para o desenvolvimento do projeto, foi definido um stack tecnológico utilizando ferramentas open-source, de modo a garantir flexibilidade, transparência e ausência de custos de licenças:

- **Python** – linguagem principal de programação e análise de dados.
- **Jupyter Notebook** – ambiente para análise exploratória e documentação interativa.
- **Visual Studio Code (VS Code)** - ambiente de desenvolvimento integrado (IDE) utilizado para escrever, testar e executar código.
- **Pandas e NumPy** – bibliotecas para manipulação e transformação de dados.
- **Matplotlib e Seaborn** – bibliotecas para visualização para análise exploratória.
- **Scikit-learn** - biblioteca de Machine Learning utilizada para modelação e avaliação.

Adicionalmente, para o Projeto 2 (fase de Deployment), está previsto o uso de ferramentas complementares de integração e visualização, nomeadamente Apache Airflow, PostgreSQL e Apache Superset.

## 2.2.2. Requisitos, Premissas e Restrições

Esta secção identifica os requisitos técnicos e organizacionais necessários à execução do projeto, as premissas assumidas e as restrições que podem afetar o desenvolvimento do mesmo.

- **Requisitos:**

1. O conjunto de dados “Screen Time vs Mental Wellness Survey – 2025” deve estar devidamente carregado e validado quanto à sua integridade e consistência.
2. Todos os ambientes de desenvolvimento (Jupyter Notebook e VS Code) devem estar configurados com as bibliotecas essenciais para análise e modelação em Python.
3. Os resultados obtidos (análises, métricas e modelos) devem ser totalmente reproduzíveis, com scripts documentados e armazenados num repositório comum.

- **Premissas:**

1. Assume-se que as respostas do survey refletem, de forma fidedigna, os comportamentos e estados de bem-estar dos indivíduos inquiridos.
2. Pressupõe-se que os dados foram recolhidos e anonimizados de acordo com princípios éticos, não existindo informação sensível identificável.

- **Restrições:**

1. O trabalho deve ser desenvolvido dentro do prazo académico estabelecido, não sendo possível prolongar as fases de execução.
2. O projeto será implementado apenas com ferramentas open-source, limitando a utilização de software comercial.
3. Os recursos computacionais disponíveis são limitados a máquinas pessoais e plataformas gratuitas, restringindo o uso de algoritmos que exijam processamento intensivo.



**4.** O dataset é estático e não poderá ser expandido com novas observações ou variáveis externas.

### 2.2.3. Riscos e Contingências

Durante o desenvolvimento do projeto, podem ocorrer eventos que comprometam o cumprimento de objetivos, a qualidade dos resultados ou o calendário definido. Esta secção identifica os principais riscos e as respetivas contingências.

**Risco 1:** Qualidade ou quantidade de dados insuficientes

- **Descrição:** Existe a possibilidade de o dataset conter valores incoerentes, outliers ou variáveis com baixa relevância para a análise, ou de o mesmo conter um volume de dados insuficiente.
- **Contingência:** Realizar uma verificação detalhada da qualidade dos dados durante a fase de Data Understanding, aplicando técnicas de limpeza e imputação, bem como a exclusão de colunas redundantes.

**Risco 2:** Desempenho insatisfatório dos modelos preditivos

- **Descrição:** Os modelos de regressão ou classificação podem apresentar baixa capacidade de generalização devido à variabilidade limitada dos dados ou à presença de ruído.
- **Contingência:** Testar múltiplos algoritmos (por exemplo, Regressão Linear, Random Forest, SVM) e ajustar hiperparâmetros através de cross-validation. Documentar comparativamente as métricas e selecionar o modelo mais robusto.

**Risco 3:** Limitações computacionais

- **Descrição:** A execução de tarefas de treino e validação pode ser afetada por restrições de hardware local.
- **Contingência:** Otimizar scripts através de amostragem de dados ou redução de dimensionalidade.

**Risco 4:** Atrasos na execução das fases do projeto



- **Descrição:** A coordenação entre membros do grupo ou imprevistos técnicos podem originar atrasos no cronograma.
- **Contingência:** Definir uma metodologia de trabalho de modo a acompanhar o progresso do projeto em grupo. Reajustar tarefas conforme necessário para garantir o sucesso do projeto.

#### Risco 5: Interpretação incorreta dos resultados

- **Descrição:** A análise dos outputs dos modelos pode levar a conclusões enviesadas ou mal fundamentadas do ponto de vista clínico.
- **Contingência:** Validar os resultados com base em métricas objetivas e garantir uma leitura crítica dos padrões encontrados, destacando limitações e possíveis enviesamentos na secção de avaliação.

#### 2.2.4. Terminologia

Esta secção define os principais termos e conceitos utilizados ao longo do projeto, garantindo uma compreensão comum entre os intervenientes técnicos e não técnicos.

- **Glossário de Negócio:**

- **Tempo de Ecrã (Screen Time):** Quantidade média de horas diárias que um indivíduo passa a utilizar dispositivos digitais, incluindo computador, smartphone e tablet.
- **Bem-Estar Mental (Mental Wellness):** Estado psicológico global de um indivíduo, refletido em valores como stress, qualidade do sono, produtividade e equilíbrio emocional.
- **Qualidade do Sono (Sleep Quality):** Medida subjetiva e quantitativa da qualidade do descanso noturno, avaliada numa escala de 1 a 5.
- **Nível de Stress (Stress Level):** Indicador do grau de tensão psicológica ou emocional, medido numa escala de 0 a 10.
- **Produtividade (Productivity):** Grau de desempenho percebido no trabalho ou nas atividades diárias, numa escala de 0 a 100.

- **Clínica de Cuidados Mentais:** Organização interessada em compreender os padrões digitais e psicológicos dos seus pacientes, recorrendo a dados para apoiar intervenções personalizadas.
- **Glossário de Mineração de Dados:**
  - **Data Mining:** Processo de descoberta de padrões e relações relevantes em grandes volumes de dados, através de métodos estatísticos e de Machine Learning.
  - **Dataset:** Conjunto estruturado de dados, composto por observações (linhas) e variáveis (colunas) utilizados para análise e modelação.
  - **Feature:** Atributo individual de um dataset que contém informação sobre uma característica dos indivíduos analisados.
  - **Target:** Variável que se pretende prever ou explicar através de modelos de Machine Learning (ex.: sleep\_quality\_1\_5 ou stress\_level\_0\_10).
  - **Pré-processamento de Dados (Data preprocessing):** Etapa que envolve limpeza, transformação e normalização dos dados antes da aplicação de algoritmos.
  - **Modelo Preditivo (Predictive Model):** Algoritmo treinado para estimar o valor de uma variável alvo com base em variáveis explicativas.
  - **Avaliação do Modelo (Model Evaluation):** Processo de medição da performance de um modelo através de métricas como RMSE, R<sup>2</sup>, F1-Score ou Acurácia.

### 2.2.5. Custos e Benefícios

Esta secção apresenta uma estimativa qualitativa dos custos associados ao desenvolvimento do projeto e dos benefícios esperados, tanto diretos como indiretos, para a organização (Clínica de Cuidados Mentais) e para o contexto académico.

- **Custos:**
  - **Custos do projeto:** O projeto será desenvolvido no âmbito académico, não implicando custos financeiros diretos. No entanto, podem ser considerados custos implícitos relacionados com:
    - **Mão de obra académica:** Esforço e tempo dedicados pelos estudantes à análise, implementação e documentação do projeto.

- **Recursos de computação:** utilização de computadores pessoais e possíveis instâncias em nuvem gratuitas.
- **Infraestrutura de armazenamento:** espaço para guardar datasets, scripts e modelos em repositórios locais ou partilhados.
- **Manutenção e compatibilidade de software:** necessidade de atualização das bibliotecas Python e dos ambientes de desenvolvimento.
- **Limitações de hardware:** eventual lentidão no processamento ou treino de modelos devido ao baixo poder computacional.
- **Benefícios:**
  - **Benefícios Diretos:**
    - **Apoio à tomada de decisão clínica:** o projeto permitirá identificar relações entre tempo de ecrã e bem-estar mental, auxiliando a clínica a adotar medidas preventivas e personalizadas.
    - **Desenvolvimento de modelos preditivos úteis:** possibilitará prever fatores de risco, como má qualidade de sono ou níveis elevados de stress.
    - **Melhoria da gestão de dados:** introduz práticas estruturadas de análise e tratamento de dados dentro de um contexto de saúde mental.
  - **Benefícios Indiretos:**
    - **Aprofundamento académico:** reforço das competências técnicas dos estudantes nas áreas abordadas pelo projeto.
    - **Valor científico:** o projeto contribui para a compreensão dos efeitos do comportamento digital no bem-estar psicológico.
    - **Potencial de escalabilidade:** os resultados e os modelos desenvolvidos poderão ser integrados em diversos sistemas, permitindo uma análise contínua.

### 2.3. Determinação das Metas de Mineração de Dados

Esta seção traduz os objetivos de negócio em metas técnicas e específicas para o projeto, alinhadas com as necessidades da clínica e o conteúdo presente no dataset.



Irão ser aplicadas diferentes técnicas de Machine Learning para extrair conhecimento e apoiar as decisões da clínica com base em dados concretos que tragam benefícios para a mesma.

### 2.3.1. Metas de Mineração de Dados

- **Meta 1: Modelo de regressão para a previsão da qualidade do sono.**
  - Descrição: O grupo de trabalho irá desenvolver um modelo de regressão para prever a variável “sleep\_quality\_1\_5”. Para chegar a esta previsão iremos utilizar variáveis as variáveis disponíveis no dataset relacionadas com os hábitos dos cidadãos, as suas características pessoais e os fatores de bem-estar.
  - Objetivo: Identificar de que modo os hábitos diários em conjunto com as características pessoais e os fatores de sanidade influenciam a qualidade do sono, prevendo padrões que afetam negativamente este aspeto.
  - Métricas de sucesso:  $R^2 > 0.75$ .
  - Benefícios esperados para o negócio: Fornecer apoio à clínica na deteção precoce de pacientes com má qualidade de sono e promover procedimentos para evitar estas situações.
- **Meta 2: Modelo de classificação para a identificação de indivíduos com alto nível de stress**
  - Descrição: Criar um modelo de classificação para identificar previamente pacientes com alto nível de stress com base nas variáveis relativas à utilização de ecrãs, qualidade de sono entre outros hábitos.
  - Objetivo: Distinguir e identificar perfis de risco elevado de stress e fornecer informações valiosas que permitam à clínica planear intervenções preventivas e benéficas.
  - Métricas de sucesso: Accuracy  $\geq 85\%$
  - Benefícios esperados para o negócio: Reduzir o número de casos de stress elevado através da aplicação de estratégias direcionadas especificamente aos hábitos de cada paciente.
- **Meta 3: Modelo de clustering para a segmentação**
  - Descrição: Irão ser aplicados algoritmos de clustering para agrupar os indivíduos de acordo com algumas variáveis que a equipa de trabalho



considerou importantes na análise dos mesmos (screen\_time\_hours, leisure\_screen\_hours, stress\_level\_0\_10, sleep\_quality\_1\_5 e mental\_wellness\_index\_0\_100).

- Objetivo: Identificar e agrupar grupos de indivíduos com padrões semelhantes.
- Métricas de sucesso: Interpretabilidade dos clusters.
- Benefícios esperados para o negócio: Permitir à clínica definir planos direcionados especificamente para indivíduos com determinados padrões.

### 2.3.2. Critérios de Sucesso Técnico

O sucesso técnico será avaliado de acordo com os seguintes critérios:

- Os modelos referidos anteriormente deverão atingir as métricas mínimas estabelecidas para cada tipo de tarefa ( $R^2$ , Accuracy, Nível de interpretabilidade).
- As variáveis utilizadas e os resultados obtidos devem ser facilmente interpretáveis e coerentes com o contexto do negócio, fornecendo insights valiosos para a clínica e que ajudem a atingir os objetivos pré-estabelecidos.
- Os resultados devem ser reproduutíveis, com código bem documentado e datasets limpos e prontos a serem analisados.
- Os modelos que a equipa de trabalho achar mais robustos irão ser utilizados nas próximas fases e no Projeto 2.



## 2.4. Produção do Plano do Projeto

### 2.4.1. Plano do Projeto

#### Projeto 1

##### 1. Business Understanding – Semana 13/10/2025

###### 1.1. Análise de Requisitos

- **Descrição:** Rever o enunciado do projeto para compreender os objetivos de negócio, o contexto da "Clínica de Cuidados Mentais" e os critérios de sucesso.
- **Entregável:** Secção "Business Understanding" do relatório final.

###### 1.2. Definição de Metas de Data Mining

- **Descrição:** Traduzir os objetivos de negócio em metas técnicas claras: uma de regressão (predict sleep\_quality), uma de classificação (identificar high\_stress) e uma de clustering (segmentar utilizadores).
- **Entregável:** Definição formal das metas no relatório.

###### 1.3. Criação do Plano de Projeto

- **Descrição:** Desenvolver e validar este cronograma detalhado, alocando recursos e definindo as ferramentas a serem utilizadas.
- **Entregável:** Documento de plano de projeto.

##### 2. Data Understanding & 3. Data Preparation - Semana de 27/10/2025

#### 2. Data Understanding

##### 2.1. Coleta e Carga dos Dados

- **Descrição:** Carregar o dataset ScreenTime vs MentalWellness.csv no ambiente de análise.
- **Entregável:** Script de carregamento de dados.



## 2.2. Análise Exploratória (EDA)

- **Descrição:** Realizar uma análise estatística e visual para compreender as distribuições, correlações e identificar anomalias. Criar um dicionário de dados.
- **Entregável:** Notebook de EDA com visualizações e Dicionário de Dados.

## 2.3. Verificação de Qualidade

- **Descrição:** Identificar e documentar valores em falta, inconsistências ou erros nos dados.
- **Entregável:** Relatório de qualidade de dados na secção "Data Understanding".

# 3. Data Preparation

## 3.1. Limpeza e Transformação

- **Descrição:** Aplicar técnicas para tratar os problemas de qualidade identificados (ex: imputação). Codificar variáveis categóricas (gender, occupation) e escalar as numéricas.
- **Entregável:** Script de pré-processamento de dados.

## 3.2. Engenharia de Características

- **Descrição:** Criar variáveis relevantes, como o alvo binário high\_stress a partir de stress\_level\_0\_10.
- **Entregável:** Dataset final, limpo e preparado para a modelação.

# 4. Modeling & 5. Evaluation - Semana de 24/11/2025

## 4. Modeling

### 4.1. Divisão dos Dados



- **Descrição:** Separar o dataset preparado em conjuntos de treino e teste para garantir uma avaliação robusta dos modelos.
- **Entregável:** Script com a divisão de dados.

#### 4.2. Treino dos Modelos

- **Descrição:** Implementar e treinar os modelos selecionados para as tarefas de regressão, classificação e clustering.
- **Entregável:** Scripts de treino para cada modelo.

### 5. Evaluation

#### 5.1. Avaliação de Desempenho

- **Descrição:** Calcular e analisar as métricas de avaliação apropriadas para cada modelo.
- **Entregável:** Tabela comparativa de desempenho dos modelos.

#### 5.2. Seleção e Finalização

- **Descrição:** Comparar os resultados, selecionar os modelos com melhor desempenho para cada tarefa e guardá-los (serializar) para uso no Projeto 2.
- **Entregável:** Ficheiros dos modelos finais.

#### 5.3. Redação do Relatório (P1)

- **Descrição:** Consolidar todas as fases, metodologias, resultados e conclusões no relatório final do Projeto 1.
- **Entregável:** Relatório final do Projeto 1.

## Projeto - Screen Time vs Mental Wellness

Fase	Início	Fim	Setembro	Outubro	Novembro
Business Understanding	15/09/2025	17/10/2025			
Data Understanding	20/10/2025	24/10/2025			
Data Preparation	27/10/2025	31/10/2025			
Modelling & Evaluating	03/11/2025	28/11/2025			

Figura 1 – Gantt Chart das fases do projeto.

### 2.4.2. Avaliação Inicial de Ferramentas e Técnicas

#### Definição do Stack Tecnológico

Para a execução deste projeto, foi selecionado um conjunto de ferramentas (*stack tecnológico*) inteiramente baseado em soluções *open-source*. Esta abordagem garante a flexibilidade, a ausência de custos de licenciamento e o controlo total sobre o ambiente de desenvolvimento e produção, alinhando-se com as melhores práticas da indústria de dados.

As ferramentas foram escolhidas para cobrir todas as fases do ciclo de vida do projeto, desde a análise exploratória até à implementação final do sistema de Business Intelligence, conforme o modelo CRISP-DM.

#### *Fases 1 a 3: Business Understanding, Data Understanding e Data Preparation*

Nestas fases iniciais, o foco está na exploração, limpeza e transformação dos dados. As seguintes ferramentas serão utilizadas para análise interativa e preparação do dataset:

- **Linguagem de Programação: Python**
  - **Função:** Linguagem principal para o desenvolvimento de todo o projeto, escolhida pela sua sintaxe clara, robustez e pelo seu vasto ecossistema de bibliotecas especializadas em análise de dados e Machine Learning.
- **Ambiente de Desenvolvimento: Jupyter Notebooks e Visual Studio Code (VS Code)**
  - **Função:** O **Jupyter** será utilizado para a análise exploratória de dados (EDA), permitindo a execução interativa de código, visualização de resultados imediatos e documentação do processo. O **VS Code** servirá



como o ambiente de desenvolvimento integrado (IDE) principal para a escrita e organização dos scripts de pré-processamento e modelagem.

- **Bibliotecas de Manipulação de Dados: Pandas e NumPy**
  - **Função:** O **Pandas** será a ferramenta central para carregar, limpar, transformar e analisar os dados estruturados do ficheiro .csv. O **NumPy** fornecerá a base para operações numéricas e computação vetorial de alta performance.
- **Biblioteca de Visualização: Matplotlib (com Seaborn)**
  - **Função:** Essencial para a fase de *Data Understanding*, o **Matplotlib** e a sua interface de alto nível **Seaborn** serão utilizados para criar as visualizações estatísticas (histogramas, diagramas de dispersão, mapas de calor de correlação) necessárias para compreender a natureza e a qualidade dos dados.

### ***Fases 4 e 5: Modeling e Evaluation***

Nesta etapa, o foco é a construção, treino e avaliação dos modelos de Machine Learning.

- **Biblioteca de Machine Learning: Scikit-learn**
  - **Função:** Será a principal biblioteca para a implementação dos modelos. Utilizaremos os seus módulos para:
    - **Pré-processamento:** Escalonamento de variáveis e codificação de categorias.
    - **Modelagem:** Treino dos modelos de regressão, classificação e clustering.
    - **Avaliação:** Cálculo das métricas de desempenho (RMSE, R<sup>2</sup>, F1-Score, Coeficiente de Silhueta) e validação dos modelos.
    - **Finalização:** Serialização (exportação) dos modelos finais para uso posterior na fase de *Deployment*.