



UNIVERSIDADE  
LUSÓFONA

# Utilização de Inteligência Artificial no estudo dos comportamentos aditivos e dependências

## Trabalho Final de curso

1ª Entrega Intercalar

Afonso Amaral, 22207489, Informática de Gestão

João Carmona, 22301968, Informática de Gestão

**Orientador:** Iolanda Velho

**Co-orientador:** Diogo Morais

**Entidade Externa:** IPSS Ares do Pinhal

[www.ulusofona.pt](http://www.ulusofona.pt)

Departamento de Engenharia Informática e Sistemas de Informação

Universidade Lusófona, Centro Universitário de Lisboa

07/11/2024

---

## **Direitos de cópia**

Utilização de Inteligência Artificial no combate aos comportamentos aditivos e dependências, Copyright de Afonso Rego Monteiro Megre do Amaral e João Tiago Carmona Universidade Lusófona.

A Escola de Comunicação, Arquitetura, Artes e Tecnologias da Informação (ECATI) e a Universidade Lusófona têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

## Resumo

Os comportamentos toxicodependentes e as suas dependências são condições que afetam a capacidade de uma pessoa controlar o uso de substâncias, este tipo de comportamentos pode comprometer a saúde física, mental e as relações sociais do indivíduo. Cada caso tem causas e manifestações únicas, envolvendo fatores biológicos, emocionais e sociais. Por isso o programa de redução de riscos exige uma abordagem cuidada e especializada, centrada na recuperação e no equilíbrio das várias áreas da vida de cada pessoa.

Segundo os dados mais recentes, a mortalidade relacionada com o consumo de droga continua a crescer em Portugal. Em 2021 registaram-se, pelo terceiro ano consecutivo, mais overdoses no território nacional, mais concretamente 74. Os números dispararam 45% face a 2020, atingindo o valor mais alto desde 2009. Em 2020, estiveram em tratamento 23.932 utentes, com problemas relacionados com o uso de drogas, no ambulatório da rede pública. Os dados salientam que, dos 3.236 utentes que iniciaram tratamento, 1.538 eram readmitidos e 1.698 eram novos utentes. [1]

A Associação Ares do Pinhal (AAP) é uma Instituição Particular de Solidariedade Social (IPSS) fundada em 1986, cujo principal foco é a inclusão social de pessoas em situação de vulnerabilidade, como os toxicodependentes e os sem-abrigo. Desde a sua criação, a associação tem desenvolvido diversas respostas terapêuticas e de reinserção, incluindo a implementação de Comunidades Terapêuticas, programas de redução de danos, iniciativas de reintegração social e uma sala de consumo, para prevenir consumo a céu aberto e tentar reduzir o número de overdoses. A Ares do Pinhal tem vindo a destacar-se pela gestão de programas inovadores, como o Programa de Substituição em Baixo Limiar de Exigência, que inclui a utilização de metadona para dependentes de opioides, e o apoio psicossocial em centros de acolhimento. [2]

A Ares do Pinhal desempenha um papel vital nesta resposta, oferecendo tratamento e apoio contínuo a pessoas com dependências, através de uma rede de cuidados que visa tanto a recuperação como a reintegração social. A associação segue uma filosofia de "menos dependência, mais cidadania", procurando ajudar os utentes a reconquistar a sua autonomia e dignidade. [2]

A parceria entre a associação e a Universidade Lusófona de Humanidades e Tecnologias mais precisamente o Departamento de Engenharia Informática e Sistemas de Informação (DEISI), tem como objetivo realizar um estudo aprofundado sobre comportamentos aditivos e dependência, com foco na aplicação de tecnologias de Inteligência Artificial (IA) para o estudo e análise destes comportamentos. Esta colaboração visa não apenas implementar soluções inovadoras no estudo dos comportamentos aditivos, mas também avaliar e demonstrar a eficácia dessas tecnologias como ferramentas de intervenção, reforçando o papel da IA como um recurso valioso no tratamento e na redução de comportamentos aditivos.

O presente projeto surge no âmbito do "Trabalho Final de Curso" (TFC) e tem como objetivo exibir todo o conhecimento adquirido ao longo da Licenciatura em Informática de Gestão (LIG), inserida no plano curricular do terceiro ano, orientado pela professora Iolanda Velho e Diogo Morais na unidade curricular anual da Universidade Lusófona de Humanidades e Tecnologias (ULHT).

**Palavras-chave:** Ares do Pinhal; Toxicodependência; Comportamentos aditivo e dependências; Data Mining; Machine Learning; Inteligência Artificial.

---

# Abstract

Drug-dependent behaviors and addictions are conditions that impair a person's ability to control substance use. Such behaviors can jeopardize an individual's physical health, mental well-being, and social relationships. Each case has unique causes and manifestations, influenced by biological, emotional, and social factors. As a result, treatments require a careful, specialized approach, focused on recovery and balance across various aspects of each person's life.

According to the latest data, drug-related mortality continues to rise in Portugal. In 2021, the country recorded an increase in overdoses for the third consecutive year, with a total of 74 cases—representing a 45% surge compared to 2020 and the highest number since 2009. In 2020, 23,932 patients were treated on an outpatient basis for drug-related issues in the public health network. Of the 3,236 individuals who began treatment, 1,538 were readmitted, and 1,698 were new patients.

Ares do Pinhal Association is a Private Institution of Social Solidarity (IPSS) founded in 1986, focused on the social inclusion of vulnerable individuals, such as drug addicts and the homeless. Since its inception, the association has developed a range of therapeutic and reintegration responses, including the establishment of Therapeutic Communities, harm reduction programs, social reintegration initiatives, and a supervised consumption room to prevent open-air drug use and reduce the number of overdoses. Ares do Pinhal has stood out for its management of innovative programs, such as the Low-Threshold Substitution Program, which includes the use of methadone for opioid dependents, along with psychosocial support at shelter centers.

Ares do Pinhal plays a crucial role in this support system, providing ongoing treatment and assistance to people with addictions through a network of care aimed at both recovery and social reintegration. The association follows a philosophy of "less dependency, more citizenship," seeking to help clients regain their autonomy and dignity.

The partnership between the center and Universidade Lusófona de Humanidades e Tecnologias aims to conduct an in-depth study on addictive behaviors and dependency, with a focus on applying Artificial Intelligence (AI) technologies to the treatment of these issues. This collaboration seeks not only to implement innovative solutions to combat addiction but also to evaluate and demonstrate the effectiveness of these technologies as intervention tools, reinforcing AI's role as a valuable resource in treating and reducing addictive behaviors.

This project part of the Final Course Project (TFC) and aims to showcase all the knowledge acquired during the Bachelor's in Management Informatics (LIG), included in the third-year curriculum and overseen by professors Iolanda Velho and Diogo Morais within the annual course unit at Universidade Lusófona de Humanidades e Tecnologias (ULHT).

**Keywords:** Ares do Pinhal; Drug Addiction; Addictive Behaviors and Dependencies; Data Mining; Machine Learning; Artificial Intelligence.

# Índice

Resumo.....	iv
Abstract .....	v
Índice.....	vi
Lista de Figuras.....	viii
Lista de Siglas .....	ix
1 Introdução aos problemas .....	1
1.1 Classificação do Abandono .....	3
1.1.1 Introdução .....	3
1.1.2 Enquadramento.....	3
1.1.3 Motivação e Identificação do Problema.....	4
1.1.4 Objetivos.....	4
1.2 Análise de Clustering e segmentação .....	4
1.2.1 Introdução .....	4
1.2.2 Enquadramento.....	5
1.2.3 Motivação e Identificação do Problema.....	5
1.2.4 Objetivos.....	5
1.3 Estrutura do documento.....	6
2 Pertinência e Viabilidade .....	7
2.1 Pertinência.....	7
2.2 Viabilidade .....	8
3 Conceitos Fundamentais.....	9
3.1 Modelos e Algoritmos Relevantes .....	10
3.2 Tecnologias e Ferramentas Utilizadas .....	13
4 Estado da Arte .....	15
4.1 Estado da Arte.....	15
4.1.1 <i>Predicting Drug Addiction Using Multimodal Data Fusion and Machine Learning Techniques</i> .....	15
4.1.2 How Machine Learning is Used to Study Addiction in Digital Healthcare: A Systematic Review.....	15
4.1.3 Pontos não abordados.....	16
5 Solução Proposta.....	17

---

5.1	Introdução .....	17
5.2	Proposta de inovação e mais-valias.....	18
5.3	Metodologia.....	19
5.4	Recolha e Descrição dos Dados .....	19
5.5	Pré-processamento dos dados .....	21
5.6	Análise Exploratória dos Dados .....	23
6	Método e Planeamento .....	24
	Bibliografia .....	27
7	Anexos.....	28

## Lista de Figuras

<i>Figura 1 – Logótipo da Instituição Particular de Solidariedade Social Ares do Pinhal</i>	2
<i>Figura 2 - Carrinha de distribuição e toma de metadona da AAP</i>	3
<i>Figura 3 - Entrada do Edifício do Serviço de Apoio Integrado (SAI)</i>	4
<i>Figura 4 – Sala de Consumo (fonte: <a href="https://aresdopinhhal.pt/wp/projects/">https://aresdopinhhal.pt/wp/projects/</a>)</i>	5
<i>Figura 5 - Conceitos fundamentais para a realização do trabalho final de curso</i>	9
<i>Figura 6 - Fórmula de Regressão Logística</i>	10
<i>Figura 7 - Confusion Matrix</i>	11
<i>Figura 8 - Logótipo da aplicação Docker</i>	13
<i>Figura 9 - Logótipo da aplicação PostgreSQL</i>	13
<i>Figura 10 - Logótipo da aplicação DBeaver</i>	13
<i>Figura 11 - Bibliotecas Python usadas para Machine Learning</i>	14
<i>Figura 12 - Imagem do Diagrama Organizacional da Metodologia</i>	19
<i>Figura 13 - Estrutura Organizacional de Pré Processamento de Dados</i>	22
<i>Figura 14 - Work Breakdown Structure</i>	25
<i>Figura 15 - Gráfico de Gantt</i>	25
<i>Figura 16 - Tabela com milstones</i>	26
<i>Figura 17 - Imagem do Esquema da Base de Ddos da AAP</i>	28
<i>Figura 18 - Imagem 1 da Tabela de Colunas da Base de Dados</i>	28
<i>Figura 19 - Imagem 3 da Tabela de Colunas da Base de Dados</i>	28
<i>Figura 20 - Imagem 3 da Tabela de Colunas da Base de Dados</i>	28



---

## Lista de Siglas

LIG	Licenciatura em Informática de Gestão
TFC	Trabalho Final de Curso
IPSS	Instituição Particular de Solidariedade Social
AAP	Associação Ares do Pinhal
PSBLE	Programa de Substituição em Baixo Limiar de Exigência
SAI	Serviço de Apoio Integrado
CAD	Comportamentos aditivo e dependências
CML	Câmara Municipal de Lisboa
ML	Machine Learning
IA	Inteligência Artificial
OCR	Optical Character Recognition
RNN	Recurrent Neural Networks
LSTM	Long Short Term Memory
NLP	Natural Language Processing
RF	Random Forests
SVM	Support Vector Machines
DL	Deep Learning
KNN	K-Nearest-Neighbors



# **1 Introdução aos problemas**

Os comportamentos aditivos e o consumo de substâncias psicoativas constituem desafios urgentes no campo da saúde pública, com repercussões sociais e económicas significativas. Estas questões afetam diretamente pessoas em situações de grande vulnerabilidade, exigindo abordagens específicas e adaptadas às suas realidades. Para enfrentar estes desafios, a Ares do Pinhal, uma organização de referência em Portugal, implementa programas inovadores e orientados para a redução de danos, a inclusão social e o acesso a cuidados adequados.

Entre os programas geridos pela Ares do Pinhal, destacam-se dois que desempenham um papel essencial no apoio a estas populações: o Programa de Substituição em Baixo Limiar de Exigência (PSBLE) e o Serviço de Apoio Integrado (SAI).

O PSBLE é uma iniciativa que procura proporcionar suporte a pessoas com comportamentos aditivos que enfrentam barreiras no acesso a tratamentos convencionais. Este programa adota uma abordagem de baixo limiar de exigência, que facilita o acesso de utentes em situações de grande vulnerabilidade, oferecendo alternativas como a substituição de substâncias de alto risco por outras mais seguras. Esta metodologia tem como objetivo principal reduzir danos e promover a estabilidade, criando um ambiente que incentive a adesão a programas de apoio.[\[3\]](#)

No âmbito do Programa de Substituição em baixo Limiar de Exigência, um dos principais desafios identificados é o abandono por parte dos utentes, um fenómeno multifacetado. Este abandono pode assumir dimensões negativas, como o regresso ao consumo descontrolado ou à exclusão social, mas também positivas, como a transição para programas de tratamento mais estruturados ou formalizados. A compreensão dos fatores que contribuem para estes diferentes tipos de abandono, utilizando modelos preditivos, é fundamental para identificar perfis de risco e implementar estratégias que promovam maior adesão e transições bem-sucedidas.

O SAI é um serviço focado em oferecer uma resposta integrada e personalizada a pessoas com múltiplas necessidades, abrangendo desde questões relacionadas com consumo de substâncias até desafios sociais e de saúde física ou mental. Este programa procura responder à complexidade das realidades enfrentadas pelos utentes, promovendo a inclusão e a dignidade através de uma abordagem holística e centrada no indivíduo.

No Serviço de Apoio Integrado, a diversidade dos utentes atendidos constitui um elemento-chave para a eficácia das intervenções. Dado que os indivíduos apresentam diferentes padrões de consumo, condições de saúde e realidades sociais, torna-se essencial compreender esta heterogeneidade. A segmentação da população, utilizando técnicas como clustering, permite identificar grupos com necessidades semelhantes, facilitando a formulação de políticas personalizadas e a alocação eficiente de recursos. Esta abordagem garante um acesso mais equitativo a serviços adequados e eficazes, promovendo a segurança e a dignidade dos utentes.

A utilização de técnicas de Inteligência Artificial (IA), em particular de modelos de Machine Learning (ML), apresenta-se como uma abordagem inovadora e promissora para responder a estes desafios. Estas ferramentas permitem explorar padrões complexos nos dados, identificar fatores determinantes no comportamento dos utentes e propor soluções preditivas e personalizadas. Apesar do impacto crescente da IA em diversas áreas da saúde, a sua aplicação no campo das dependências e no apoio a populações vulneráveis permanece relativamente pouco explorada. Este projeto pretende, assim, contribuir para preencher esta lacuna, demonstrando o potencial destas tecnologias para melhorar a eficácia dos programas de intervenção e otimizar a alocação de recursos, com benefícios significativos para as populações apoiadas.

Neste trabalho, as problemáticas do abandono no PSBLE e da segmentação de utentes no SAI, ambos programas geridos pela Ares do Pinhal, serão analisadas com o objetivo de desenvolver soluções baseadas em dados que ampliem a eficácia e o alcance desses programas de apoio.



*Figura 1 – Logótipo da Instituição Particular de Solidariedade Social Ares do Pinhal*

*(fonte: <https://aresdopinhal.pt/wp/>)*

Neste TFC, que é composto por dois alunos, tem como objetivo abordar dois problemas distintos, de elevada relevância no contexto dos comportamentos aditivos e dependências:

- **Classificação do Abandono**
- **Análise de Clustering e Segmentação**

Em seguida, será apresentada uma descrição detalhada de cada problema, destacando a sua relevância, os métodos a utilizar e os objetivos a alcançar neste TFC. Este trabalho tem como propósito desenvolver soluções fundamentadas em evidências que sustentem decisões estratégicas e promovam a eficácia e a evolução contínua dos programas de apoio analisados.

## 1.1 Classificação do Abandono

### 1.1.1 Introdução

O consumo de substâncias psicoativas e as dependências químicas representam uma das maiores dificuldades de saúde pública em Portugal e no mundo. Para além dos comprovados efeitos infelizes em termos de saúde, tanto física como mental, estas condições causam consequências sociais de profunda gravidade: a exclusão, desemprego e incremento da criminalidade associada. Neste contexto, os programas de redução de danos surgem como uma solução de ‘salvação’ para mitigar riscos associados ao consumo das substâncias, sobretudo em populações vulneráveis que, por múltiplas razões, não conseguem inserir-se em programas mais estruturados de tratamento.

O Programa de Substituição em Baixo Limiar de Exigência (PSBLE) é uma iniciativa da IPSS Ares do Pinhal que procura assistir os precursores de drogas opióides que, por razões diversas, não estão em capacidade ou não querem aceder às estruturas convencionais de tratamento. Trata-se, assim, de um programa de ambulatório que providencia o apoio da área da saúde pública e psicossocial, com inclusão da entrega diária de metadona, para mitigação dos riscos pessoais, sociais e de saúde associados ao consumo de substâncias ilícitas e lícitas. [\[3\]](#)



Figura 2 - Carrinha de distribuição e toma de metadona da AAP

### 1.1.2 Enquadramento

No âmbito do PSBLE, o abandono constitui um dos desafios mais relevantes para a eficácia do tratamento. O abandono ocorre quando um utente não comparece a uma sessão durante mais de sete dias consecutivos sem origem aparente. Na unidade descrita, no entanto, esta ausência é formalmente notificada como baixa do programa após um mês – 30 dias contínuos de ausência consecutiva. Essa longa ausência costuma ser um indicativo de múltiplos fatores, tais como recaídas no uso, motivos alheios à vontade, ou entraves ao acesso ao programa, sejam estes de origem logística ou pessoal. Seja como for, a alta descontinuidade do tratamento paralisa o processo, deixando o utente em estado de maior vulnerabilidade, o que aumenta o risco de deterioração da saúde, exclusão social e, em situações extremas, até mesmo de exposição a perigo de vida.

### **1.1.3 Motivação e Identificação do Problema**

Reconhecer e compreender o problema do abandono é essencial para desenvolver intervenções mais eficazes, capazes de manter os utentes no programa e maximizar os benefícios da sua participação.

### **1.1.4 Objetivos**

Iremos analisar as saídas dos utentes do programa, com o objetivo de compreender e classificar estas transições como positivas ou negativas, avaliando os fatores que contribuem para cada tipo de desfecho.

Uma saída positiva ocorre quando um utente é transferido para instituições que oferecem suporte mais intensivo, como comunidades terapêuticas ou programas de reabilitação, com o objetivo de parar completamente o consumo de substâncias. Estas transições são planeadas e representam um avanço no tratamento, indicando que o utente está preparado para a próxima etapa no processo de recuperação.

Por outro lado, as saídas negativas implicam que o utente abandona o programa sem qualquer plano de continuidade. Estas situações ocorrem frequentemente quando os utentes faltam às tomas diárias de metadona ou não demonstram motivação para mudar. Como resultado, há um aumento significativo do risco de recaídas e da continuação do consumo, colocando tanto o próprio utente como a comunidade em situações de maior vulnerabilidade e risco.

## **1.2 Análise de Clustering e segmentação**

### **1.2.1 Introdução**

O Serviço de Apoio Integrado (SAI) é uma iniciativa desenvolvida no âmbito das políticas de Redução de Riscos e Minimização de Danos, criada para atender exclusivamente pessoas com comportamentos aditivos e dependências (CAD). Promovido pela Câmara Municipal de Lisboa e gerido pela Associação Ares do Pinhal, o projeto destaca-se por promover dignidade e inclusão, disponibilizando um espaço seguro e supervisionado para o consumo de substâncias, além de oferecer serviços de saúde, higiene e lazer. [\[4\]](#)

Para melhorar a eficácia do programa, a aplicação de Análise de Clustering e Segmentação surge como uma abordagem inovadora. Este método permite analisar e agrupar os utentes de acordo com suas características e necessidades, viabilizando intervenções mais eficazes e personalizadas.



*Figura 3 - Entrada do Edifício do Serviço de Apoio Integrado (SAI)*

### 1.2.2 Enquadramento

O SAI foi estruturado com base no **Decreto-Lei 183/2001**, posicionando-se como uma resposta integrada às necessidades de uma população vulnerável. Através de serviços como o Programa de Consumo Vigiado, rastreios de saúde e encaminhamento para redes de apoio, o SAI contribui para a proteção da saúde pública e individual, reduzindo riscos associados ao consumo de substâncias. [\[4\]](#)

Complementarmente, oferece espaços para higiene, alimentação e atividades de lazer, promovendo bem-estar e dignidade entre os utentes. Essas iniciativas têm como objetivo central minimizar os danos relacionados ao consumo, enquanto se procura criar condições para uma eventual transição para programas de reabilitação.

### 1.2.3 Motivação e Identificação do Problema

O SAI desempenha um papel fundamental no apoio às pessoas com comportamentos aditivos e dependências, respondendo às suas necessidades de forma estruturada e multidisciplinar. No entanto, devido à diversidade nos perfis dos utentes, torna-se essencial compreender melhor as suas características e necessidades específicas.

Sem uma análise detalhada e segmentada dos utentes, é difícil adaptar os serviços de maneira a maximizar a eficácia das intervenções e oferecer suporte personalizado. A implementação de metodologias como o Clustering Segmentation permitiria identificar padrões e perfis, facilitando uma resposta mais ajustada às diferentes realidades dos utentes.

### 1.2.4 Objetivos

Um dos principais objetivos deste trabalho é implementar a metodologia de Análise de Clustering e Segmentação no âmbito do Serviço de Apoio Integrado (SAI), de modo a identificar e classificar os utentes em perfis semelhantes, permitindo uma compreensão mais aprofundada das suas necessidades e características e viabilizando uma abordagem personalizada e eficaz. Paralelamente, procuramos otimizar as intervenções realizadas pelo SAI, ajustando os recursos disponíveis de acordo com os padrões de consumo, necessidades específicas e utilização dos serviços identificados, de forma a atender com maior precisão cada grupo, promovendo maior eficiência na gestão dos serviços.



Figura 4 – Sala de Consumo

### **1.3 Estrutura do documento**

Na **secção 1, Introdução aos Problemas**, apresentamos os desafios relacionados com os comportamentos aditivos e o consumo de substâncias psicoativas, destacando o impacto significativo que estes fenómenos têm na saúde pública e na inclusão social. Esta secção introduz os dois problemas centrais que serão abordados ao longo do projeto: o abandono dos utentes nos programas de apoio, com foco no Programa de Substituição em Baixo Limiar de Exigência (PSBLE), e a necessidade de segmentação no Serviço de Apoio Integrado (SAI) para uma melhor personalização das intervenções.

Adicionalmente, são descritas a motivação para o desenvolvimento deste trabalho e os objetivos a alcançar. A motivação surge da necessidade de estudar e ajudar os programas de apoio geridos pela Ares do Pinhal, utilizando técnicas de Inteligência Artificial para compreender melhor o comportamento dos utentes e propor soluções baseadas em dados.

Na **secção 2, Pertinência e Viabilidade**, discutimos a relevância e o impacto positivo esperado da integração de Inteligência Artificial (IA) no tratamento de dependências. Analisamos, ainda, a viabilidade técnica, económica e social do projeto, destacando os benefícios esperados para os programas de apoio analisados.

Na **secção 3, Conceitos Fundamentais**, apresentamos os conceitos teóricos essenciais para o projeto, como Base de Dados, Machine Learning e técnicas de Data Mining. Nesta secção, descrevemos também as ferramentas e tecnologias utilizadas, justificando a escolha com base na adequação às necessidades do projeto.

Na **secção 4, Estado da Arte**, analisamos estudos anteriores sobre a aplicação de Inteligência Artificial (IA) no combate às dependências químicas, identificando lacunas nas abordagens existentes. Por fim, destacamos como o projeto pretende propor inovações e melhorias face às metodologias tradicionais.

Na **secção 5, Solução Proposta**, especificamos e modelamos os problemas abordados, propondo soluções baseadas em IA. Descrevemos, de forma detalhada, a metodologia aplicada, incluindo a recolha e análise de dados, os modelos utilizados e as estratégias definidas para a implementação das soluções.

Na **secção 6, Método e Planeamento**, apresentamos o cronograma detalhado do projeto, com recurso a ferramentas como o Diagrama de Gantt. Nesta secção, detalhamos as principais etapas do trabalho, desde as reuniões iniciais até à entrega do relatório intercalar, assegurando o acompanhamento estruturado e sistemático das tarefas.



## 2 Pertinência e Viabilidade

### 2.1 Pertinência

O projeto proposto, que integra tecnologias IA no tratamento de comportamentos aditivos e dependência química, pretende alcançar um impacto significativo e contribui diretamente para a resolução do problema identificado. A aplicação da IA tem um grande papel na resolução de grandes desafios e problemas da área como, a personalização do tratamento, a detecção precoce de comportamentos de risco e o apoio à reintegração social. Estes aspetos foram destacados como limitações das abordagens terapêuticas tradicionais.

#### Validação por Terceiros

A aplicação de tecnologias avançadas, como a Inteligência Artificial (IA), no tratamento de dependências tem sido analisada e testada por especialistas na área da saúde mental. Estudos recentes indicam que a IA pode aumentar a precisão no diagnóstico e na personalização das intervenções, além de melhorar os resultados a longo prazo. Por exemplo, uma investigação chamada “Aplicação da Inteligência Artificial em transtornos mentais: uma revisão sistemática”[\[5\]](#) destaca que **"a IA tem o potencial de transformar a maneira como compreendemos, diagnosticamos e tratamos os transtornos mentais, oferecendo uma perspetiva inovadora diante das limitações dos métodos tradicionais"**

#### Impacto Positivo Esperado

Os benefícios esperados do projeto incluem:

1. **Melhorias no conhecimento e suporte à decisão:** Intervenções mais direcionadas e adaptadas às necessidades específicas de cada utente.
2. **Redução de recaídas e abandonos:** Detecção precoce de comportamentos de risco através de ferramentas preditivas.
3. **Apoio mais robusto à reintegração social:** Planeamento individualizado baseado em análises de dados.

## **2.2 Viabilidade**

A viabilidade do projeto foi avaliada com base em critérios técnicos, económicos e sociais, mostrando que a solução proposta pode ser implementada e sustentada com sucesso.

### **1. Viabilidade Técnica**

O projeto utiliza tecnologias disponíveis no mercado, como algoritmos de machine learning (ML) e sistemas de apoio à decisão, que podem ser integrados nas ferramentas existentes da AAP. A base de dados da AAP é adequada para sustentar os modelos preditivos, facilitando a implementação técnica.

Além disso, a viabilidade técnica é reforçada pela **cooperação entre a Ares do Pinhal e o DEISI**, estabelecida no âmbito de um Trabalho Final de Curso (TFC) realizado no último ano letivo. Esta parceria permite o desenvolvimento contínuo de soluções técnicas alinhadas às necessidades da associação.

Adicionalmente, a **formalização iminente de um protocolo de cooperação científica entre a Ares do Pinhal e o DEISI** reforça a sustentabilidade económica do projeto. Este acordo formaliza a colaboração entre as instituições e assegura o suporte necessário para a implementação de soluções práticas.

### **2. Viabilidade Económica**

Uma análise de custo-benefício inicial demonstra que o projeto é sustentável. A implementação de IA pode reduzir custos operacionais a longo prazo ao otimizar recursos humanos e melhorar a eficácia dos tratamentos

### **3. Viabilidade Social**

Após uma visita ao estabelecimento e reuniões com o Dr. Hugo Faria e a Dr. Cláudia Pereira foi nos informados que havia uma aceitação elevada por parte das partes interessadas, incluindo utentes e profissionais de saúde e que este trabalho poderia vir a resolver problemas na comunidade do bairro do Casal Ventoso e de toda a área Metropolitana de Lisboa. A proposta atende às expectativas de personalização do tratamento e apoio à recuperação social, elementos valorizados por parte dos profissionais que trabalham na AAP e objetivo de beneficiar os utentes.

Além disso, a **colaboração contínua com os responsáveis da Ares do Pinhal**, que vão fornecer os dados e o apoio necessário no decorrer deste ano letivo, assegura que a solução seja ajustada às reais necessidades da associação e da comunidade. O **compromisso com a adequação do estudo e dos modelos desenvolvidos** é resultado direto de um diálogo constante, garantindo que o projeto responda efetivamente às prioridades identificadas pela Ares do Pinhal.

Em suma, com base na análise de pertinência e viabilidade, conclui-se que o projeto apresenta uma elevada capacidade de implementação técnica, económica e social, e que tem potencial transformador para enfrentar os desafios atuais no tratamento de toxicodependência. O impacto positivo esperado reforça a importância de avançar com esta proposta como um modelo sustentável e reaplicável.

### 3 Conceitos Fundamentais

Para o desenvolvimento deste trabalho, é essencial compreender e aplicar conceitos fundamentais como bases de dados, linguagens de programação, Data Mining, Machine Learning e Ciência de Dados. Estes elementos são interdependentes e indispensáveis para alcançar os objetivos propostos.

Primeiramente, é necessário utilizar uma base de dados para armazenar e gerir os dados que servirão como ponto de partida do projeto. No caso específico deste trabalho, utilizaremos a base de dados fornecida pela Ares do Pinhal, que será importada para o ambiente de análise. As bases de dados são essenciais para organizar e estruturar a informação que será trabalhada e analisada nas fases seguintes.

De seguida, a linguagem de programação desempenha um papel central no processamento e manipulação dos dados. Para este trabalho, usaremos Python, uma das linguagens mais populares e versáteis em projetos de análise de dados e Inteligência Artificial. Através da biblioteca Pandas, NumPy e Matplotlib extrairemos os dados, faremos a sua transformação e análise inicial, o que permitirá efetuar operações como limpeza, reestruturação e visualização da informação de forma eficiente.

Com os dados preparados, aplicaremos técnicas de Data Mining para identificar padrões, relações e tendências importantes. Esta fase é essencial para explorar os dados de forma sistemática, transformando a informação bruta em conhecimento útil. Adicionalmente, usaremos métodos de Machine Learning (ML) assim como uma das bibliotecas de Python, Scikit Learn, que possibilitam a criação de modelos preditivos e a descoberta de insights baseados nos padrões identificados nos dados.

Estas etapas interligadas constituem a base metodológica para a realização deste trabalho, desde a importação e preparação da base de dados até à aplicação de técnicas avançadas que visam extrair valor e gerar soluções práticas.

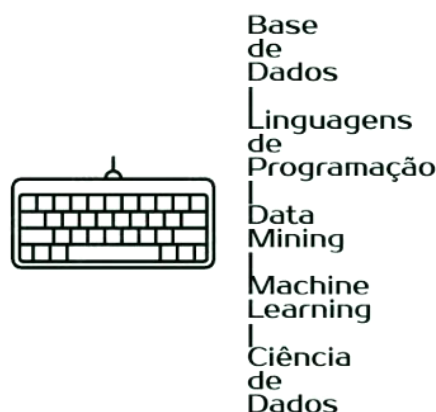


Figura 5 - Conceitos fundamentais para a realização do trabalho final de curso

### 3.1 Modelos e Algoritmos Relevantes

Neste trabalho, serão utilizadas várias tecnologias, bibliotecas e algoritmos para processar e analisar os dados. A seleção de cada ferramenta e técnica foi fundamentada na sua capacidade de lidar com os requisitos específicos de cada fase do projeto. As etapas e os recursos escolhidos são descritos abaixo:

#### 3.1.1 Classificação do abandono

Nesta etapa, os modelos selecionados inicialmente para o desenvolvimento e análise são o Logistic Regression e o K-Nearest Neighbors (KNN), assim como as métricas de para cimentar e garantir que os dados estão corretos na previsão e classificação do abandono.

- **Regressão Logística**

A Regressão Logística é um modelo linear, o que significa que ela modela a relação entre variáveis independentes (features) e a variável dependente (alvo) usando uma função logística (sigmoide). A fórmula básica é:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

Figura 6 - Fórmula de Regressão Logística

A Regressão Logística é especialmente útil neste caso, onde o objetivo é prever o comportamento de abandono em dois possíveis resultados:

- Abandono positivo (1): Indicando que o indivíduo reduziu o consumo e avançou para tratamentos.
- Abandono negativo (0): Indicando que o indivíduo voltou a consumir e abandonou o programa.

O valor obtido através do modelo de Regressão Logística é uma probabilidade (um valor entre 0 e 1). Essa probabilidade permite definir um limiar de classificação:

- Por exemplo, um limiar padrão é 0.5:
  - $P(y=1|X) > 0.5$  → Classificado como 1 (abandono positivo).
  - $P(y=1|X) \leq 0.5$  → Classificado como 0 (abandono negativo).

- **K-Nearest-Neighbors**

O modelo **K-Nearest Neighbors (KNN)** foi escolhido para a etapa de classificação pela sua flexibilidade e capacidade de lidar com problemas tanto de classificação quanto de regressão.

O funcionamento do KNN baseia-se na análise da proximidade entre os dados. O modelo prevê a classe de uma nova instância com base nas classes das k instâncias mais próximas no espaço de atributos. O valor de k é um parâmetro crucial: valores baixos de k podem levar ao overfitting, enquanto valores altos podem causar underfitting. O valor ideal será ajustado através de validação cruzada para garantir um equilíbrio entre esses dois extremos.

- **Métricas de classificação**

Ao finalizar a análise com o modelo de **Regressão Logística**, a avaliação de desempenho é crucial para garantir que o modelo seja eficiente na previsão do comportamento de abandono. As principais métricas aplicadas incluem:

### Confusion Matrix

We grade and tally our predictions in a confusion matrix.

TN = True Negative

FN = False Negative

TP = True Positive

FP = False Positive

	Predicted ✗	Predicted ✓
Actual ✗	TN	FP
Actual ✓	FN	TP

Figura 7 - Confusion Matrix

#### 1. Accuracy

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Soma a previsão correta dividida pela soma de todas as previsões.

#### 2. Recall

$$Recall = \frac{TP}{TP + FN}$$

Mede a habilidade do modelo para detetar ocorrências de uma classe.

#### 3. Precision

$$Precision = \frac{TP}{TP + FP}$$

Mede a habilidade de um modelo para evitar falsos alarmes para uma classe.

#### 4. F1 Score

$$F1 = \frac{Precision \times Recall}{Precision + Recall}$$

Faz uma combinação das métricas Precision e Recall.

#### 3.1.2 Análise de Clustering e Segmentação

O Serviço de Apoio Integrado (SAI) tem como objetivo oferecer uma resposta personalizada a pessoas com múltiplas necessidades, desde problemas relacionados com consumo de substâncias até questões de saúde física, mental e desafios sociais. A diversidade dos utentes atendidos exige uma compreensão aprofundada da sua heterogeneidade, o que pode ser facilitado por técnicas como o K-Means e o Principal Component Analysis (PCA).

- **K-Means**

O K-Means é uma técnica de clustering que permite organizar objetos em grupos cujos membros são similares de alguma forma. Os passos básicos incluem escolher o número de clusters (k), inicializar centróides aleatoriamente, atribuir pontos ao centróide mais próximo, recalculando os centróides e repetir até a convergência ou atingir o número máximo de iterações. Este processo pode trazer benefícios diretos ao SAI:

- Segmentação de Utes: Identificar grupos homogêneos de utentes com base em dados como por exemplo: tipos de substâncias consumidas, frequência de consumo, condições de saúde física e mental e nível de apoio social.
- Personalização das Intervenções: Os clusters criados permitem desenhar intervenções adaptadas a cada grupo, como programas de tratamento específicos ou estratégias de reintegração social.
- Eficiência Operacional: O algoritmo é rápido e pode ser executado diversas vezes com diferentes inicializações para otimizar os resultados, usando o método do cotovelo para determinar o número ideal de clusters (k).

- **PCA**

O PCA é uma técnica de redução de dimensionalidade que permite simplificar dados complexos, mantendo as informações mais relevantes. É particularmente útil no contexto do SAI:

- Redução de Variáveis: Os utentes apresentam múltiplas características (idade, padrões de consumo, condições de saúde). O PCA transforma estas variáveis em componentes principais, condensando a informação mais significativa.
- Preparação para K-Means: O PCA simplifica os dados, melhorando a separação geométrica entre grupos, o que torna o K-Means mais eficaz e os clusters mais precisos.

- Visualização de Dados: Permite criar gráficos bidimensionais ou tridimensionais para explorar os padrões e identificar relações entre os utentes.
- Identificação de Casos Atípicos (Outliers): Utentes que não se encaixam em nenhum cluster podem ser rapidamente identificados, sinalizando a necessidade de intervenções individualizadas.

### **Aplicação e Métricas de Avaliação**

O uso integrado do K-Means e do PCA no Serviço de Apoio Integrado facilita a compreensão e gestão da diversidade dos utentes. Estas técnicas permitem segmentar a população de forma precisa, criando intervenções focadas, eficazes e alinhadas com os princípios de inclusão e dignidade que o serviço promove.

## **3.2 Tecnologias e Ferramentas Utilizadas**

### **1ª Fase: Configuração e Gestão da Base de Dados**

Para a 1ª fase, foi diagnosticado que é necessário implementar uma solução eficiente para o armazenamento e gestão de dados. Será utilizado o PostgreSQL, um sistema bastante utilizado, que permite armazenar e consultar grandes volumes de dados de forma estruturada. A interface com a base de dados será feita através do DBeaver, uma ferramenta gráfica que facilita a gestão, consulta e manipulação dos dados.

Adicionalmente, foi utilizado o Docker para criar não só um container com a base de dados da associação como também um ambiente de desenvolvimento isolado garantindo que todas as dependências e configurações necessárias para o PostgreSQL e o DBeaver funcionem corretamente.



*Figura 9 - Logótipo da aplicação PostgreSQL*



*Figura 10 - Logótipo da aplicação DBeaver*



*Figura 8 - Logótipo da aplicação Docker*

## 2ª Fase: Processamento e Análise dos Dados

Na 2ª fase, a análise dos dados será feita utilizando bibliotecas e ferramentas orientadas à linguagem Python. Esta etapa envolve a utilização de Pandas, uma biblioteca poderosa para manipulação e análise de dados. O Pandas será utilizado para extrair, transformar e estruturar os dados provenientes da base de dados PostgreSQL.

Além disso, as bibliotecas NumPy, Matplotlib, Pandas e Scikit Learn serão implementadas para apoiar a análise e visualização dos dados:

- NumPy: Esta biblioteca será utilizada para realizar cálculos matemáticos e manipular arrays multidimensionais. A sua eficiência no processamento de grandes volumes de dados numéricos torna-a indispensável para operações matemáticas avançadas e manipulação de matrizes.
- Matplotlib: Será implementada para criar gráficos e representações visuais dos dados. A visualização dos dados é crucial para a interpretação dos resultados e para a comunicação das descobertas de forma clara e acessível.
- Pandas: Será utilizada para manipulação e análise de dados estruturados. A capacidade do Pandas de lidar com estruturas como DataFrames e Séries é essencial para a organização, limpeza e transformação dos dados. Esta biblioteca será particularmente útil para realizar operações como filtragem, agrupamento e agregação de dados.
- Scikit Learn: Esta biblioteca será implementada para construir e avaliar os modelos de machine learning utilizados neste trabalho, incluindo a Logistic Regression, o K-Nearest Neighbors (KNN), e outros algoritmos, caso necessário.



Figura 11 - Bibliotecas Python usadas para Machine Learning



## 4 Estado da Arte

### 4.1 Estado da Arte

O uso de tecnologias de inteligência artificial (IA) para abordar questões relacionadas à dependência química está a começar a ganhar atenção significativa em estudos recentes. Este avanço reflete a crescente necessidade de soluções inovadoras para lidar com a complexidade desse problema, que afeta não apenas os indivíduos diretamente envolvidos, mas também suas famílias, comunidades e o sistema de saúde pública.

Neste contexto, estudos recentes têm explorado diferentes aplicações de IA para enfrentar os desafios da dependência química. Em seguida, descrevemos os dois artigos que encontramos, nos quais destacaremos e detalharemos as abordagens apresentadas.

#### 4.1.1 *Predicting Drug Addiction Using Multimodal Data Fusion and Machine Learning Techniques*

Um exemplo que encontramos é o trabalho de Dhanya S [6] com o acompanhamento de estudantes do departamento de Computer Science and Engineering, LBSITW, Trivandrum (2024), que desenvolveram um modelo para prever dependência de drogas utilizando uma abordagem de fusão de dados multimodais. O estudo combinou análises de **questionários comportamentais, exames de sangue e análise de voz** utilizando tecnologia como:

- Optical Character Recognition (OCR)
- Recurrent Neural Networks (RNN)
- Modelos Long Short Term Memory (LSTM)

A pesquisa destacou-se pela capacidade de integrar diferentes fontes de dados para melhorar a precisão diagnóstica e oferecer suporte personalizado através de chatbots utilizando Processamento de Linguagem Natural (NLP). Este trabalho demonstra como a combinação de tecnologias avançadas pode aumentar a precisão no diagnóstico e facilitar intervenções personalizadas.

#### 4.1.2 *How Machine Learning is Used to Study Addiction in Digital Healthcare: A Systematic Review*

O estudo de Chhetri et al. (2023) [7] apresenta uma revisão sistemática sobre a aplicação de técnicas de Machine Learning (ML) no estudo de transtornos e de dependência. A pesquisa analisa 26 estudos que utilizam métodos baseados em ML, como:

1. **Supervised Learning**
  - Random Forests (RF)
  - Support Vector Machines (SVM)
2. **Unsupervised Learning**
  - K-Means
3. **Deep Learning (DL)**

Aplicados em dados de imagens cerebrais, questionários comportamentais e sinais elétricos cerebrais.

O estudo destaca que ML tem se mostrado eficaz na identificação precoce de comportamentos de risco e no diagnóstico de transtornos relacionados ao abuso de substâncias, através de modelos, com uma alta precisão. Além disso, os autores salientam a importância do uso de tecnologias como:

1. Ressonância magnética funcional
2. Sinais de EEG

As anteriores mencionadas, combinadas com dados comportamentais, são capazes de classificar padrões de dependência e oferecer insights sobre a conectividade funcional cerebral.

#### **4.1.3 Pontos não abordados**

Apesar de apresentarem contribuições significativas no avanço das tecnologias aplicadas à dependência, os trabalhos de Dhanya S. et al. e Chhetri et al. possuem limitações que o projeto da AAP tenta abordar. Ambos os estudos se destacam pela inovação no uso de ML e IA para diagnóstico, com precisão elevada em métricas específicas. No entanto, limitam-se a estágios iniciais do cuidado, não abordando elementos críticos como a personalização de longo prazo ou estratégias de reintegração social dos utentes.

O projeto da AAP diferencia-se por ir além da identificação de padrões e comportamentos de risco. Ele foca na integração de tecnologias avançadas para oferecer intervenções completas e centradas no utente, incluindo:

- **Classificação de abandono:** Estudar e analisar abandonos dos utentes no tratamento de metadona.
- **Apoio Abrangente à Reintegração Social:** Desenvolve estratégias para ajudar os utentes na recuperação sustentável e redução de danos, enfrentando os desafios de reintegração no tecido social e reduzindo o risco de recaídas.
- **Uso Contextualizado de Bases de Dados Nacionais:** Diferente dos trabalhos referenciados, o projeto da AAP aproveita bases de dados nacionais para criar soluções adaptadas ao sistema de saúde pública em Portugal, garantindo maior aplicabilidade e impacto local.

Essas características tornam o projeto da AAP uma solução mais completa e inovadora, abordando lacunas identificadas nos estudos anteriores e oferecendo um modelo replicável para avanços futuros no tratamento e na prevenção da dependência química

## **5 Solução Proposta**

### **5.1 Introdução**

No caso do Programa de Substituição em Baixo Limiar de Exigência (PSBLE), a proposta envolve a implementação de um sistema de monitorização preditiva e intervenções direcionadas para prevenir o abandono dos utentes. Inicialmente, serão aplicados modelos de Machine Learning de classificação, como Logistic Regression e K-Nearest Neighbors (KNN), para identificar perfis de risco de abandono com base em dados históricos e características do utente. Esses modelos permitirão análises iniciais eficientes, enquanto a avaliação contínua irá determinar a necessidade de adotar possíveis modelos mais robustos e complexos, dependendo da performance e das especificidades dos dados. Essa abordagem possibilita intervenções proativas, promovendo a adesão ao tratamento e facilitando transições para programas mais estruturados e exigentes, mitigando o impacto do abandono e aumentando a estabilidade no tratamento.

No âmbito do Serviço de Apoio Integrado (SAI), a solução proposta envolve a aplicação de Análise de Clustering e Segmentação da população assistida em grupos semelhantes com base nas características e necessidades do utente. A segmentação permitirá estudar, analisar e compreender a população de utentes, de modo a caracterizá-los e ajudar em intervenções mais personalizadas e eficazes ao alocar recursos e ampliar o efeito das ações. Isto assegura um apoio adaptado à realidade do utente, mantendo a sua dignidade, segurança e acessibilidade ao serviço de acordo com o perfil identificado.

O trabalho representa uma mais-valia científica ao:

- Expandir o conhecimento sobre a aplicação prática de IA em contextos de saúde pública específicos.
- Demonstrar como a personalização e a integração podem melhorar os resultados no tratamento da dependência química.
- Gerar dados e modelos replicáveis para futuros estudos, contribuindo para uma abordagem global mais abrangente e eficaz no combate à dependência.

Com essas características, o projeto propõe um avanço significativo no campo, não apenas fornecendo soluções práticas, mas trazer também inovações na saúde pública.

## **5.2 Proposta de inovação e mais-valias**

A solução apresentada pelo projeto da AAP destaca-se como uma abordagem inovadora no contexto do tratamento da dependência química, incorporando elementos que a diferenciam significativamente das metodologias existentes. Abaixo são identificados os principais fatores de inovação, diferenciação, vantagens e benefícios proporcionados:

### **1. Inovação na Integração Tecnológica**

- **Uso de Inteligência Artificial (IA):** A proposta utiliza tecnologias avançadas de IA para personalizar tratamentos e prever comportamentos de risco.
- **Aplicação Prática Centrada no Utente:** Ferramentas de IA são utilizadas não apenas para análise, mas para apoiar continuamente os utentes.

### **2. Diferenciação em Relação às Abordagens Existentes**

- **Classificação e previsão de abandono de tratamento:** Em comparação com os estudos já abordados, que se limitam à identificação e classificação de transtornos, o projeto da AAP oferece um modelo integrado que cobre a luta que a associação tem em minimizar danos e classificar abandonos por parte dos utentes.
- **Foco na Sustentabilidade:** A proposta considera a viabilidade de longo prazo, utilizando ferramentas económicas e sustentáveis que reduzem os custos operacionais e promovem uma recuperação mais eficiente.
- **Alinhamento com Necessidades Locais:** A utilização de uma base de dados nacional e a consideração de fatores socioculturais tornam a solução mais relevante e eficaz no contexto da saúde pública portuguesa.

### **3. Vantagens e Benefícios da Proposta**

#### **1. Melhoria na Eficiência**

- A precisão dos algoritmos de IA permite intervenções mais rápidas e eficazes, reduzindo o tempo de resposta e aumentando a taxa de sucesso no tratamento.
- A utilização de dados robustos otimiza a alocação de recursos, diminuindo custos associados a recaídas, abandonos e internações prolongadas.

#### **2. Impacto Social Relevante**

- O objetivo de reintegração social, apoiada por ferramentas de IA, promove a inclusão de utentes e apoio no tratamento com objetivo de trazer mais humanidade e uma vida melhor a estas comunidades.
- A solução aborda fatores psicossociais que influenciam o sucesso da recuperação, promovendo uma visão de solidariedade.

### 5.3 Metodologia

A metodologia desenvolvida para implementar as soluções nos dois contextos — o PSBLE e o SAI — é composta por etapas estruturadas que guiam a identificação, análise e aplicação de estratégias eficazes. Essas etapas são baseadas em processos analíticos e ações práticas para garantir o sucesso das soluções propostas. A seguir, descrevem-se sumariamente os passos principais:

1. **Definição do Problema e Objetivos:** Compreender o contexto, as dificuldades e as metas de intervenção;
2. **Análise de Dados:** Identificar, organizar e explorar os dados relacionados aos utentes;
3. **Modelagem Preditiva e Clustering:** Aplicar técnicas de machine learning para identificar perfis de risco (PSBLE) e criar agrupamentos homogêneos (SAI);

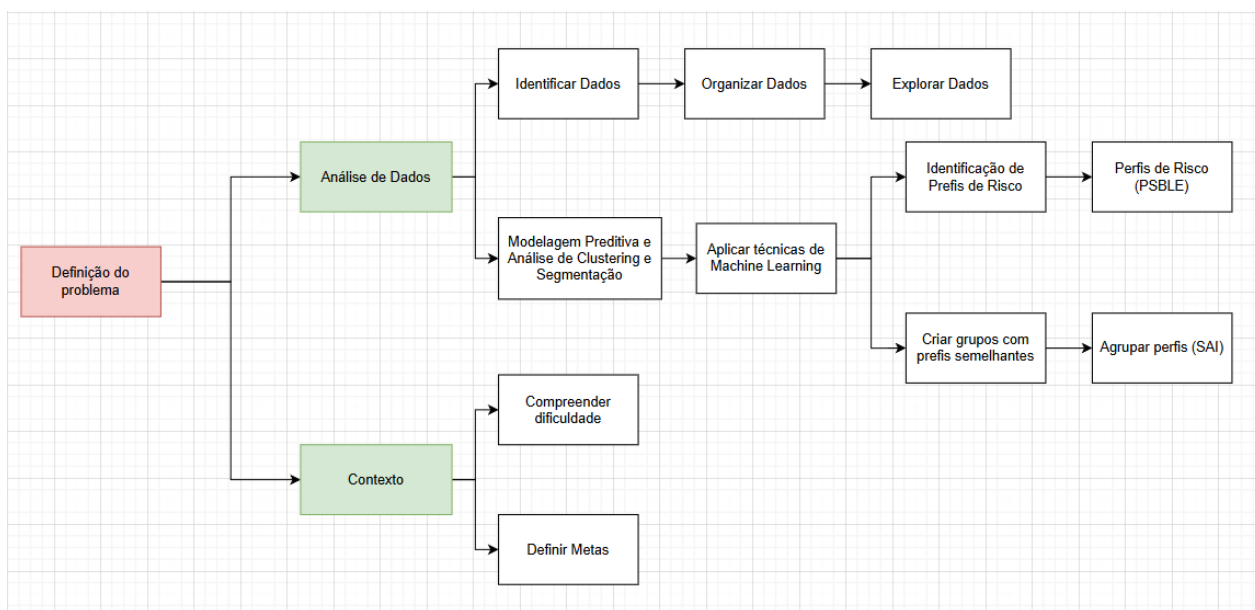


Figura 12 - Imagem do Diagrama Organizacional da Metodologia

### 5.4 Recolha e Descrição dos Dados

Os dados analisados neste trabalho foram disponibilizados pela Associação Ares do Pinhal, uma Instituição Particular de Solidariedade Social (IPSS) dedicada ao apoio de populações vulneráveis, com foco especial em comportamentos aditivos e dependências. A base de dados interna da associação é composta por **87 tabelas e 474 colunas**, distribuídas por entre as tabelas, contendo um total de **3.466.296 registos**, abrangendo informações detalhadas sobre os utentes e as intervenções realizadas nos diferentes programas de apoio.

A estrutura relacional da base de dados, gerida no sistema PostgreSQL, destaca-se pela vasta interligação entre as tabelas, permitindo cruzamentos eficientes entre variáveis relacionadas, como dados demográficos, identificação dos utentes, padrões de consumo, administração de medicação, desfechos de tratamento e problemas de saúde. Estas interconexões possibilitam

análises complexas e multivariadas, essenciais para compreender a relação entre diferentes dimensões dos serviços prestados e os resultados obtidos.

Além disso, a base de dados cobre um extenso período de atuação, oferecendo informações históricas que enriquecem o potencial analítico. A interligação entre tabelas, como as de utentes, medicação e abandono, permite rastrear trajetórias completas de tratamento e identificar padrões que possam ser úteis para a associação. Esta estrutura faz com que a base de dados seja uma fonte crucial para o desenvolvimento deste projeto.

### **1. Estrutura e Armazenamento dos Dados**

Os dados estão organizados numa base de dados relacional, gerida no sistema **PostgreSQL**, e que conseguimos aceder através da interface gráfica **DBeaver**. A estrutura das tabelas segue uma lógica relacional, permitindo um reconhecimento eficiente das múltiplas dimensões dos serviços prestados. Entre as principais tabelas, destacam-se:

- **Utentes:** Contém informações demográficas.
- **Doses:** Regista as doses diárias administradas no tratamento com metadona, incluindo ajustes.
- **Registo de faltas:** Documenta ausências e dados essenciais na recuperação pós ausência de tomas
- **Programa:** Fornece dados sobre entrada e saída dos utentes no programa, assim como o id do motivo que faz a ligação com a tabela de motivos
- **Problemas de Saúde:** Inclui registos de exames e testes de doenças infecciosas, como HIV, hepatite B, hepatite C e tuberculose.

### **2. Variáveis Principais**

- **Identificação do Utente:** Identificador único (e codificado) para proteger a privacidade.
- **Dados Demográficos:** Informações como idade, género e localização geográfica.
- **Padrões de Consumo:** Detalhes sobre as substâncias utilizadas, frequência de uso e alterações no padrão ao longo do tempo.
- **Registos de Entradas:** Datas e informações sobre o ingresso dos utentes na associação.
- **Registos de Medicação:** Informações sobre a administração de metadona e outros medicamentos.

### **3. Tipos de Dados**

- **Numéricos:** Variáveis quantitativas, como idade, frequência de comparecimento, duração do programa em dias e dosagens de medicamentos.
- **Categóricos:** Género, tipos de substâncias consumidas e presença de doenças infecciosas.
- **Texto:** Notas qualitativas, como observações sobre a administração de medicamentos e mensagens aos utentes.
- **Datas:** Registos de eventos significativos, como entrada e saída do programa ou mudanças no plano de tratamento.

Os dados refletem a operação da Associação Ares do Pinhal e oferecem uma vasta oportunidade para análises aprofundadas sobre os desafios enfrentados e os sucessos obtidos nos tratamentos. A base de dados, composta por 87 tabelas e 474 colunas, apresenta uma estrutura robusta, mas complexa, que reflete a multiplicidade de serviços e dimensões abordadas pela associação. Essa complexidade, que inclui uma grande quantidade de variáveis interligadas, é evidenciada no **Anexo**, onde apresentamos um print screen das tabelas (**Figuras 17,18 e 19**) e uma imagem do diagrama da base de dados (**Figura 16**), destacando a abrangência e riqueza das informações disponíveis.

Como é comum em bases de dados reais, esta também apresenta lacunas e inconsistências ocasionais, o que exige etapas de pré-processamento para garantir a qualidade dos dados antes de sua utilização. No entanto, essa estrutura abrangente não só facilita a compreensão do impacto das intervenções realizadas pela Ares do Pinhal, como também apoia diretamente o desenvolvimento deste projeto, fornecendo dados reais de fácil acesso e de grande valor analítico.

## **5.5 Pré-processamento dos dados**

O pré-processamento dos dados constitui uma etapa essencial em qualquer projeto de análise ou modelação preditiva. Esta fase é responsável por garantir que os dados se encontram adequadamente preparados para serem utilizados pelos algoritmos, assegurando a qualidade, consistência e representatividade necessária para obter resultados fiáveis e robustos. O principal objetivo do pré-processamento é transformar o conjunto de dados brutos em dados limpos e estruturados, eliminando ruído, inconsistências ou informações irrelevantes, e destacando as variáveis mais importantes para os modelos. As etapas realizadas no pré-processamento, conforme ilustrado na figura, são descritas em detalhe abaixo:

### **1. Exploração dos Dados**

A exploração inicial dos dados é o primeiro passo no pré-processamento. Esta etapa permite compreender a estrutura do conjunto de dados, identificar as principais características e observar padrões ou irregularidades. Esta análise exploratória é fundamental para definir o caminho a seguir nas etapas subsequentes de limpeza e transformação.

### **2. Limpeza de dados**

A limpeza dos dados é uma das fases mais importantes do pré-processamento, pois garante a eliminação de elementos que possam prejudicar os resultados. Esta etapa inclui:

- **Deteção de Duplicados:** Identificação e remoção de registos duplicados, que podem influenciar os resultados dos modelos.
- **Tratamento de Missing Values:** Lidar com valores em falta, substituindo-os por valores estimados (imputação) ou eliminando as observações, dependendo da sua relevância.
- **Identificação e Tratamento de Outliers:** Deteção de valores extremos que possam influenciar negativamente os modelos e o seu desempenho. Estes valores podem ser corrigidos ou removidos com base em análises estatísticas.

- **Scaling:** Normalização ou padronização dos dados para garantir que todas as variáveis se encontram numa escala comparável, especialmente quando utilizadas em algoritmos sensíveis à magnitude das variáveis.
- **Balancing:** Em casos de desequilíbrio nas classes a serem estudadas (por exemplo, num problema de classificação com classes desiguais), são aplicadas técnicas como oversampling ou undersampling para corrigir esta discrepância.

### 3. Engenharia de Features

Nesta fase, os dados são transformados ou enriquecidos para melhorar a performance dos modelos. As principais tarefas incluem:

- **Encoding:** Conversão de variáveis categóricas em representações numéricas (por exemplo, one-hot encoding ou label encoding), facilitando a sua utilização por algoritmos.
- **Descritizing:** Agrupamento de valores contínuos em categorias (bins) para reduzir a complexidade dos dados.
- **Criação de Novas Features:** A geração de novas features pode ajudar a capturar informações relevantes que não são diretamente observáveis nas variáveis originais.

### 4. Seleção de Variáveis

A seleção de variáveis consiste em identificar e manter apenas as variáveis mais relevantes para o problema em análise, eliminando aquelas que são redundantes ou irrelevantes. Esta etapa inclui:

- **Correlação entre Variáveis:** Identificar e tratar variáveis altamente correlacionadas, evitando redundâncias nos dados.
- **Feature Selection:** Utilização de algoritmos ou métricas que quantifiquem a relevância de cada variável para a previsão, como métodos baseados em importância ou redução de dimensionalidade.

### 5. Modelação e Reajustes

Após o pré-processamento, são realizados testes iniciais com modelos para avaliar a qualidade das variáveis selecionadas. Caso necessário, o conjunto de dados pode ser reajustado ou refinado para melhorar os resultados.

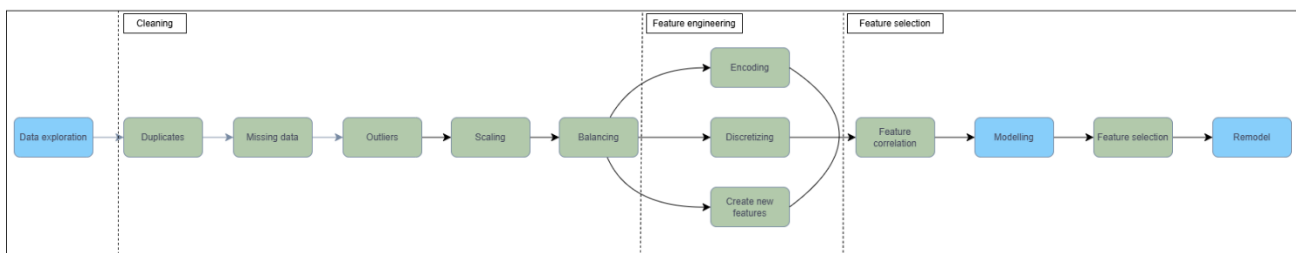


Figura 13 - Estrutura Organizacional de Pré Processamento de Dados



## **5.6 Análise Exploratória dos Dados**

Neste ponto serão apresentados os gráficos resultantes da exploração dos dados fornecidos pela Associação Ares do Pinhal. Embora a análise ainda não tenha sido iniciada, os gráficos serão desenvolvidos e incluídos em entregas futuras, permitindo uma visualização clara e detalhada das informações extraídas. Estes gráficos terão como objetivo ilustrar padrões, tendências e insights relevantes obtidos a partir da base de dados, enriquecendo a compreensão do impacto das intervenções realizadas pela associação.

## 6 Método e Planeamento

O cronograma da ordem de trabalhos foi desenvolvido com o auxílio de ferramentas de metodologia Agile, destacando-se o uso do Diagrama de Gantt (Gantt Chart) para o planeamento. Esta abordagem permitiu não só organizar as tarefas e marcos do projeto de forma estruturada, mas também integrar as expectativas temporais previstas para cada etapa do desenvolvimento. O cronograma foi ajustado para considerar fatores externos como pausas letivas, épocas de exames e responsabilidades académicas paralelas, garantindo a viabilidade prática do plano de execução.

### Etapas principais do trabalho

Na primeira etapa do projeto, foram realizadas as seguintes atividades principais:

1. **Reuniões com a orientadora:**
  - Alinhamento inicial sobre os objetivos do projeto.
  - Definição das expectativas quanto ao formato e ao conteúdo do relatório.
2. **Planeamento do relatório:**
  - Criação de um esboço inicial do conteúdo e organização das partes do relatório.
  - Elaboração de um plano de execução detalhado para orientar as próximas etapas.
3. **Reunião com a associação:**
  - Interação direta com os stakeholders envolvidos no projeto para compreender as suas necessidades e expectativas.
  - Definição de requisitos e acesso às informações relevantes, como bases de dados.
4. **Acesso e análise da base de dados:**
  - Avaliação das informações disponíveis para planear a pesquisa e a aplicação prática dos dados no desenvolvimento do projeto.
5. **Discussão e definição de objetivos:**
  - Discussão com a orientadora para definir os objetivos do projeto, alinhando-os com as métricas estabelecidas e os dados disponíveis.
6. **Pesquisa bibliográfica:**
  - Revisão de literatura científica e técnica.
  - Organização das fontes para futuras referências no relatório.
7. **Entrega do primeiro relatório completo:**
  - Compilação e documentação das informações e atividades realizadas até o momento.
  - Submissão para revisão e feedback da orientadora.

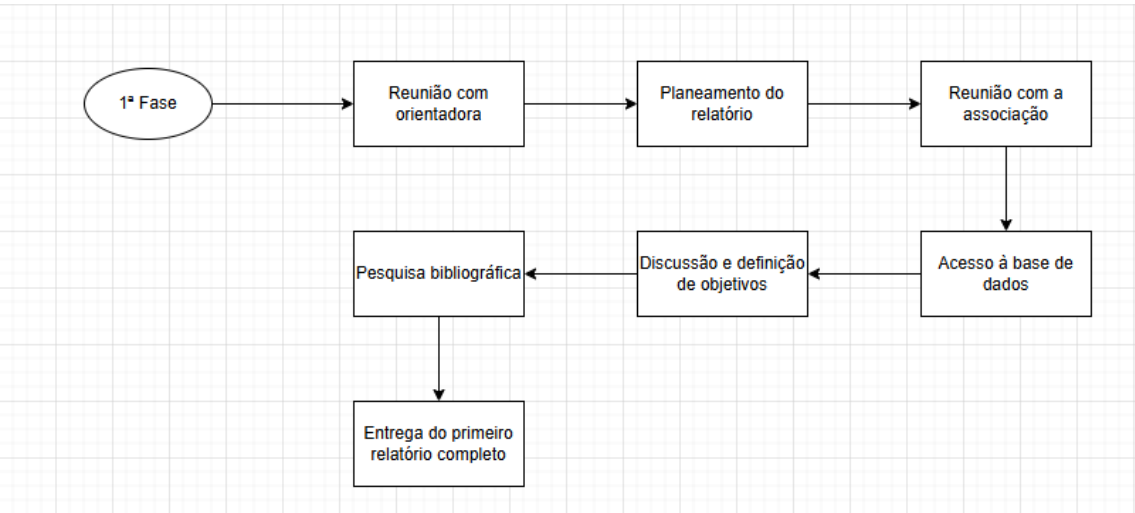


Figura 14 - Work Breakdown Structure

O diagrama seguinte mostra as tarefas ao longo do tempo com barras horizontais, indicando a sequência e dependências entre elas. Há marcos específicos como **15-10-2024** e **01-12-2024**, que representam momentos importantes como a entrega do relatório.

Esta etapa está marcada para **01-12-2024**, como indicado no Gantt e detalhado na tabela. Representa a conclusão de uma fase inicial significativa do projeto.

Este planeamento detalhado, com as tarefas estruturadas e suas dependências bem mapeadas, permite uma execução clara e organizada. As milestones funcionam como pontos de verificação importantes para medir o progresso.

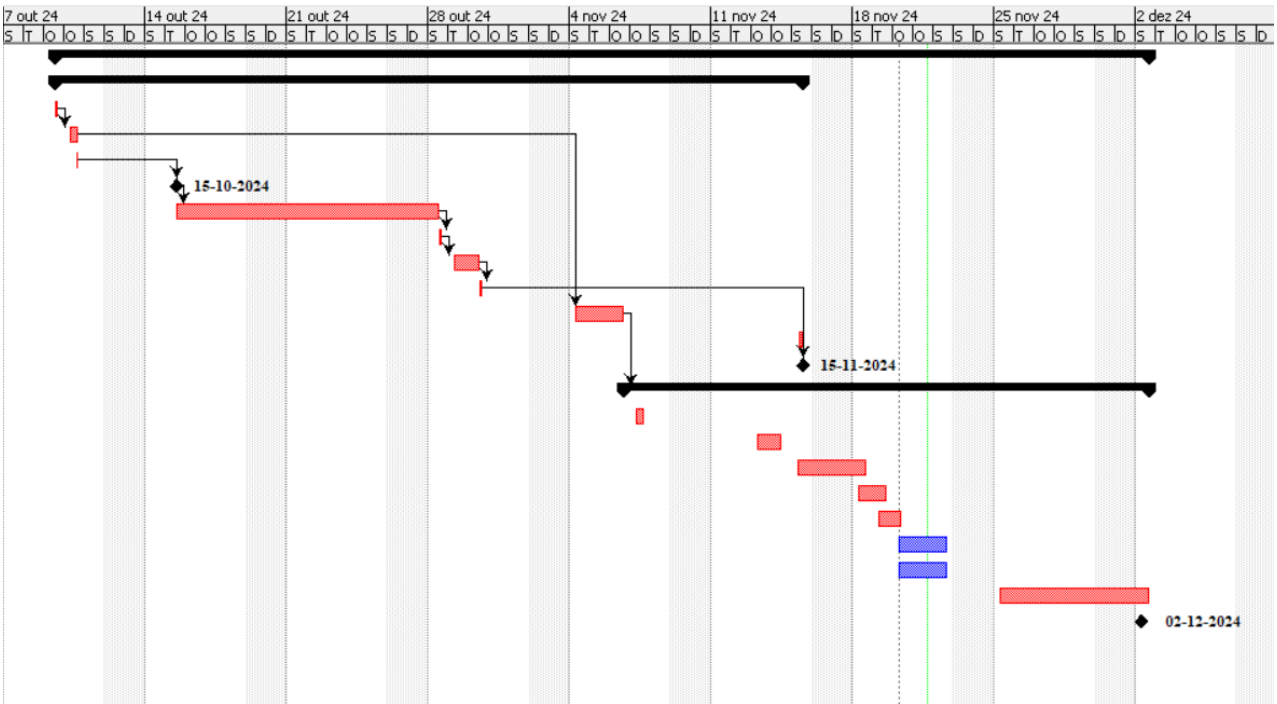


Figura 15 - Gráfico de Gantt

## Utilização de Inteligência Artificial no estudo dos comportamentos aditivos e dependências




















		Nome	Duração	Início	Fim
1		<b>TFC</b>	<b>38,25 dias</b>	<b>09-10-2024 15:00</b>	<b>02-12-2024 17:00</b>
2		<b>1ªEtapa TFC</b>	<b>26,938 dias</b>	<b>09-10-2024 15:00</b>	<b>15-11-2024 14:30</b>
3		Reunião com a Orientadora	0,25 dias	09-10-2024 15:00	09-10-2024 17:00
4		Organização da estrutura do relatório	1 dia	09-10-2024 18:00	10-10-2024 17:00
5		Reunião com Associação	0,125 dias	10-10-2024 16:00	10-10-2024 17:00
6		Acesso à base de dados	0 dias	15-10-2024 15:00	15-10-2024 15:00
7		Análise dos dados	9 dias	15-10-2024 15:00	28-10-2024 15:00
8		Definição de objetivos com orientadora	0,25 dias	28-10-2024 15:00	28-10-2024 17:00
9		Preparação de apresentação para a associação sobre os objetivos da investigação	1,75 dias	28-10-2024 17:00	30-10-2024 15:00
10		Apresentação e discussão de objetivos com associação	0,25 dias	30-10-2024 15:00	30-10-2024 17:00
11		Pesquisa Bibliográfica	3 dias	04-11-2024 8:00	06-11-2024 17:00
12		Visita à associação	0,5 dias	15-11-2024 9:30	15-11-2024 14:30
13		Definição do problema	0 dias	15-11-2024 14:30	15-11-2024 14:30
14		<b>Relatório Intercalar 1ºSemestre</b>	<b>18 dias</b>	<b>06-11-2024 17:00</b>	<b>02-12-2024 17:00</b>
15		Abstract	1 dia	06-11-2024 17:00	07-11-2024 17:00
16		Viabilidade e Pretinência	1,5 dias	13-11-2024 8:00	14-11-2024 13:00
17		Estado da Arte	2 dias	15-11-2024 8:00	18-11-2024 17:00
18		Introdução aos problemas	2 dias	16-11-2024 8:00	19-11-2024 17:00
19		Conceitos fundamentais	1,4 dias	19-11-2024 8:00	20-11-2024 11:12
20		Solução Proposta	3 dias	20-11-2024 8:00	22-11-2024 17:00
21		Conclusão	3 dias	20-11-2024 8:00	22-11-2024 17:00
22		Revisão e melhorias	6 dias	22-11-2024 17:00	02-12-2024 17:00
23		Primeira Entrega Relatório	0 dias	01-12-2024 8:00	02-12-2024 17:00

Figura 16 - Tabela com milstones

## Bibliografia

- [1] <https://www.blogtecniquitel.com/drogas-pesadas-em-portugal/>
- [2] Site Associação Ares Pinhal : <https://aresdopinhal.pt/wp/>
- [3] <https://aresdopinhal.pt/wp/wp-content/uploads/2023/01/2022-02-29-ARESDOPINHAL-FACTSHEET-PSBLE.pdf>
- [4] Sai - Serviço de apoio Integrado: <https://aresdopinhal.pt/wp/wp-content/uploads/2023/01/2022-02-15-ARESDOPINHAL-FACTSHEET.pdf>
- [5] Aplicação da Inteligência Artificial em transtornos mentais: uma revisão sistemática.(relatório)
- [6] [Dhanya, S., Kartha, P., Nair, P., Reji, R., & Suresh, S. \(2024\). Predicting Drug Addiction Using Multimodal Data Fusion and Machine Learning Techniques. International Journal of Computer Science Trends and Technology](#) (artigo)
- [7] [How Machine Learning is Used to Study Addiction in Digital Healthcare: A Systematic Review](#) (artigo site)

Projetos de Intervenção : <https://aresdopinhal.pt/wp/projects/>

30 anos a fazer da ciência uma evidência:  
[https://dependencias.pt/ficheiros/conteudos/files/DEP\\_14%20Ares%20do%20Pinhal.pdf](https://dependencias.pt/ficheiros/conteudos/files/DEP_14%20Ares%20do%20Pinhal.pdf)

Comunidade Terapeutica Ares do Pinhal: <https://aresdopinhal.pt/wp/wp-content/uploads/2023/01/2022-02-29-ARESDOPINHAL-FACTSHEET-CT.pdf>

## 7 Anexos

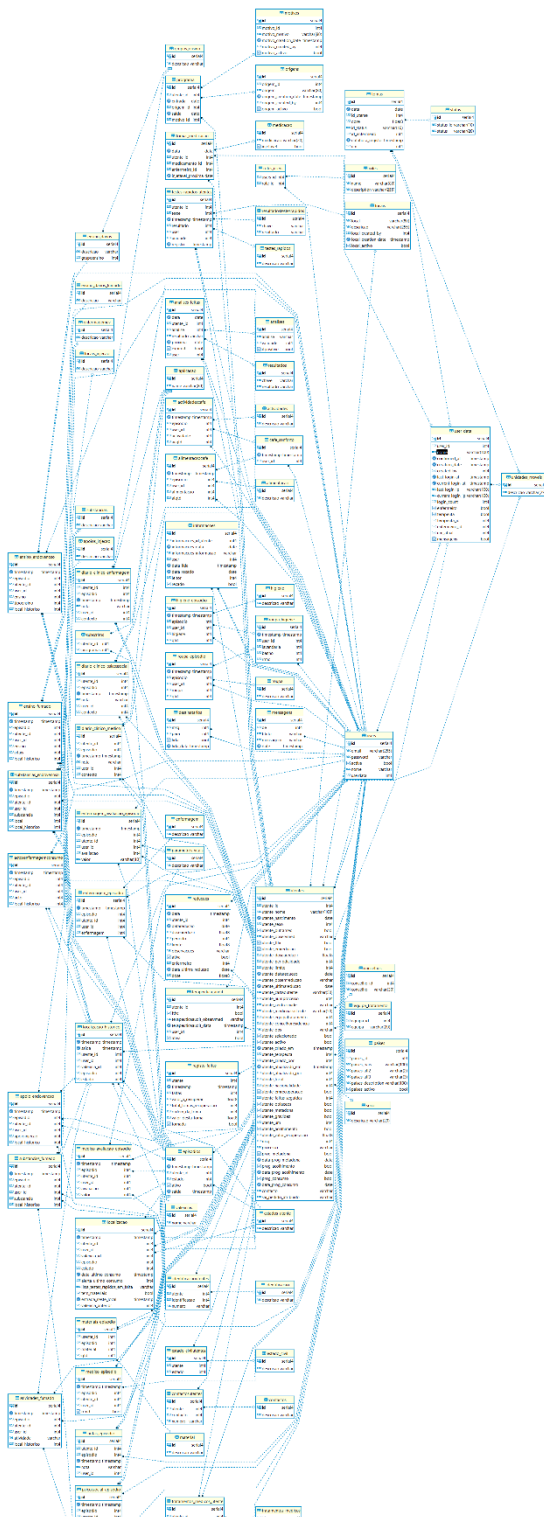


Figura 17 - Imagem do Esquema da Base de Ddos da AAP

- ▼ Tabelas
- > actividades
  - > actividadescafe
  - > actosenfermagemconsumo
  - > alembic\_version
  - > alimentacao
  - > alimentacaocafe
  - > analises
  - > analises\_feitas
  - > aplicacao
  - > apoio\_endovenoso
  - > apoios\_injecao
  - > atividades\_fumado
  - > cafe\_conforto
  - > calculo
  - > concelhos
  - > contactos
  - > contactosutentes
  - > destinatarios
  - > diario\_clinico\_enfermagem
  - > diario\_clinico\_medico
  - > diario\_clinico\_psicossocial
  - > enfermagem
  - > enfermagem\_avaliacao\_episodio
  - > enfermagem\_episodio
  - > enfermagemev

Figura 18 - Imagem 1 da Tabela de Colunas da Base de Dados

- > resultadostestesrapidos
- > roles
- > roles\_users
- > roupa
- > roupa\_episodio
- > roupa\_higiene
- > sai\_comunitaria
- > sexo
- > status
- > subscritos
- > substancias
- > substancias\_endovenoso
- > substancias\_fumado
- > terapeutica\_audit
- > testes\_rapidos
- > testes\_rapidos\_utentes
- > tomas
- > tomas\_medicao
- > tratamentos\_medicos
- > tratamentos\_medicos\_utente
- > unidades\_moveis
- > update\_consumo\_index
- > user\_data
- > users
- > utentes

Figura 19 - Imagem 3 da Tabela de Colunas da Base de Dados

- > ensino\_danos
- > ensino\_danos\_fumado
- > ensino\_endovenoso
- > ensino\_fumado
- > episodios
- > equipa\_tratamento
- > estado\_civil
- > estado\_civilutentes
- > estados\_utente
- > grupos\_ensino
- > higiene
- > higiene\_episodio
- > identificacao
- > identificacaoutentes
- > informacoes
- > locais
- > locais\_injecao
- > localizacao
- > localizacao\_historico
- > materiais\_episodio
- > material
- > medicacao
- > medico\_avaliacao\_episodio
- > medico\_episodio
- > mensagens
- > motivos
- > notas\_episodio
- > origens
- > paises
- > parametros\_vitais
- > programa
- > psicossocial\_episodio
- > reducoes

Figura 20 - Imagem 3 da Tabela de Colunas da Base de Dados