



UNIVERSIDADE
LUSÓFONA

Previsão de Tempos de Espera

Trabalho Final de Curso

1ª Entrega Intercalar

Ana Toscano, a22207260, Ciência de Dados

Nome do Orientador: Iolanda Velho

Nome do Coorientador: Sofia Naique

Entidade Externa: Centro de Medicina de Reabilitação de Alcoitão

Departamento de Engenharia Informática e Sistemas de Informação

Universidade Lusófona, Centro Universitário de Lisboa

1 de dezembro de 2024

www.ulusofona.pt

Direitos de cópia

Previsão de Tempos de Espera, Copyright de Ana Margarida Teixeira Toscano, Universidade Lusófona.

A Escola de Comunicação, Arquitectura, Artes e Tecnologias da Informação (ECATI) e a Universidade Lusófona (UL) 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 objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Este documento foi gerado com o processador (pdf/Xe/Lua)LaTeX e o modelo ULThesis (v1.0.0) [1].

Resumo

Este trabalho é desenvolvido no âmbito da unidade curricular Trabalho Final de Curso (TFC) da Licenciatura de Ciência de Dados.

A Engenheira Mariana Matos, Diretora da Unidade de Gestão de Doentes do Centro de Medicina de Reabilitação de Alcoitão (CMRA), Santa Casa da Misericórdia de Lisboa, entrou em contato com a Professora Iolanda Velho, propondo a cooperação científica na base deste TFC.

O CMRA dedica-se à reabilitação de pessoas portadoras de deficiência de predomínio físico ou multideficiência congénita, de todas as idades. No ano de 2020 recebeu 665 utentes tratados em internamento, além de ter realizado 72 859 atos terapêuticos em ambulatório. Dada a amplitude do atendimento realizado pelo CMRA e a sua procura pela excelência em reabilitação, a instituição aposta no desenvolvimento tecnológico dos seus processos.

Através da utilização de técnicas de Machine Learning para apoiar a gestão hospitalar, este projeto tem como objetivos:

- realizar a análise de dados, cedidos pelo CMRA
- desenvolver um modelo preditivo, que permita antecipar o tempo de espera de doentes que chegam ao CMRA, para internamento, ou ambulatório
- no âmbito deste modelo preditivo, caracterizar o referido tempo de espera, em função da sua patologia, e área clínica do CMRA a que se refere.

Abstract

This work is developed within the scope of the curricular unit Final Course Work (TFC) of the Bachelor's Degree in Data Science.

Engineer Mariana Matos, Director of the Patient Management Unit at the Alcoitão Rehabilitation Medicine Center (CMRA), Santa Casa da Misericórdia de Lisboa, reached out to Professor Iolanda Velho, proposing the scientific collaboration underlying this TFC.

The CMRA is dedicated to the rehabilitation of individuals with predominantly physical disabilities or congenital multiple disabilities, serving all age groups. In 2020, it received 665 inpatients and conducted 72,859 outpatient therapeutic procedures. Given the broad scope of services provided by the CMRA and its pursuit of excellence in rehabilitation, the institution invests in the technological development of its processes.

Through the use of Machine Learning techniques to support hospital management, this project aims to:

- analyze data provided by the CMRA;
- develop a predictive model that allows for anticipating the waiting time for patients arriving at the CMRA, whether for inpatient care or not;
- within the scope of this predictive model, characterize the waiting time based on the patient's pathology and the clinical area of the CMRA involved.

Índice

Resumo	2
Abstract	3
Índice	4
Lista de Figuras	6
Lista de Tabelas	7
1 - Introdução	8
2 - Pertinência e Viabilidade	10
2.1 Viabilidade	10
2.2 Pertinência	11
3 - Conceitos Fundamentais	12
3.1 Conceitos Teóricos	12
3.1.1 Machine Learning	12
3.1.2 Machine Learning na saúde	12
3.1.3 Tipos de Machine Learning	13
3.1.4 Preparação dos Dados	13
3.1.5 Métricas de Desempenho	13
3.1.6 Modelos a utilizar	15
4 - Estado da Arte	21
4.1 Artigos Analisados	21
5 - Solução Proposta	26
5.1 Dados a Utilizar	27

5.2 Ferramentas a utilizar	28
5.3 Abrangência	29
6 - Método e Planeamento	30
Bibliografia	32

Lista de Figuras

1	Centro de Medicina de Reabilitação de Alcoitão	8
2	Regressão Linear Simples	16
3	Regressão Linear Múltipla	16
4	Exemplo de uma árvore de decisão	17
5	Demonstração do funcionamento do algoritmo Random Forest	18
6	Exemplo de uma aplicação de SVM	19
7	Aplicação do algoritmo KNN	19
8	Resultado da aplicação do algoritmo K-Means	20
9	Fases propostas para a resolução do problema	27
10	Diagrama Gantt - 1º Semestre	31

Lista de Tabelas

1 Tabela de Comparação dos Estudos sobre Previsão de Tempos de Espera
 e Classificação de Pacientes 24

1 - Introdução

O Centro de Medicina e Reabilitação de Alcoitão (CMRA), Santa Casa da Misericórdia de Lisboa, fundado na década de 1950, é considerado o centro de Reabilitação mais importante do país. Durante 40 anos foi o único centro de reabilitação de Portugal tendo, desde o início, começado a apostar na educação de profissionais na área da reabilitação. Atualmente, continua a fornecer serviços exclusivos apostando em tecnologias inovadoras e na melhoria de condições de acolhimento dos utentes. [2]

O CMRA dedica-se à reabilitação pós-aguda de pacientes de todas as idades, com incapacidades físicas, motoras ou múltiplas, tanto congénitas quanto adquiridas, atendendo pessoas de todo o país. O seu principal objetivo é contribuir para o cumprimento de princípios de ética e da bioética na prestação de cuidados de saúde e na realização de investigação clínica, especialmente no campo das ciências da saúde. [3]

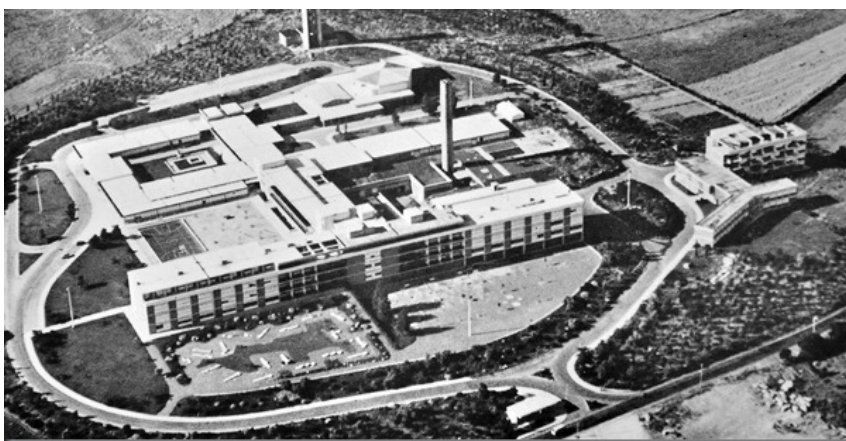


Figure 1: Centro de Medicina de Reabilitação de Alcoitão

O centro está organizado em três serviços de acordo com o grupo etário e regime de prestação de cuidados:

- **Serviço de Reabilitação de Adultos, com duas valências** – tendo a primeira 66 camas que se destinam ao internamento de utentes com lesões vertebro-medulares e outras patologias neurológicas e a segunda 68 camas destinadas ao internamento de pessoas que sofreram um acidente vascular cerebral ou traumatismo crânio-encefálico, ou que tenham sido amputadas. [4]
- **Serviço de Reabilitação Pediátrica e de Desenvolvimento** – com 16 camas, destina-se ao internamento de crianças e jovens com patologias neurológicas, osteoarticulares, medulares, entre outras. Este serviço mantém as crianças na escola, através de um professor que dá as aulas nas instalações do centro. [4]

Em função dos seus recursos, o centro enfrenta desafios críticos para responder à

crescente procura por parte dos pacientes, sendo o tempo de espera para internamento um dos principais problemas. Com uma capacidade total de apenas 150 camas (134 para adultos e 16 para pediatria), o CMRA enfrenta desafios relacionados com os tempos de espera prolongados que eventualmente poderão ter impacto na eficiência do atendimento. Além disso, a diversidade e complexidade das patologias atendidas aumenta a pressão sobre os recursos disponíveis e exige uma gestão eficaz para otimizar o atendimento. [4]

Assim, o objetivo principal deste Trabalho Final de Curso é implementar modelos de Machine Learning baseados nos dados sobre a gestão dos pacientes, fornecidos pelo centro. O objetivo principal é criar uma ferramenta que permita prever os tempos de espera para internamento e para tratamento e caracterizar os pacientes de acordo com a sua patologia. Neste sentido, com este TFC, pretende-se obter uma solução que permita apoiar a decisão, no que diz respeito aos recursos do CMRA, tanto a nível humano como de infraestruturas. Pretende-se, assim, melhorar a capacidade de gestão e a resposta à procura por parte dos pacientes. Consequentemente, espera-se obter uma solução que torne mais eficiente a distribuição de recursos, melhorando a capacidade de resposta do centro e adaptando-o à procura.

O propósito é oferecer ao centro uma visão global, mais clara e mais precisa sobre os padrões e tendências relacionados ao atendimento e à ocupação das camas de internamento. Ao promover uma análise detalhada dos dados, pretende-se que seja possível tomar decisões de forma mais informada e estratégica, o que pode ajudar a reduzir os tempos de espera.

Além disso, espera-se que a solução a propor por este TFC, aumente a autonomia do centro na gestão das suas atividades, possibilitando ajustes dinâmicos e respostas mais rápidas às necessidades dos pacientes. Ao disponibilizar informações que favoreçam uma administração proativa, o centro estará mais preparado para adaptar-se a mudanças na procura e para responder com facilidade a emergências, assegurando um atendimento de maior qualidade e eficiência.

2 - Pertinência e Viabilidade

Nos últimos anos, a aplicação Machine Learning (ML) em diferentes contextos clínicos, tem-se tornado cada vez mais valioso na otimização de recursos e de gestão de atendimento. No CMRA, onde a crescente procura e os limites existentes na sua capacidade de internamento são pontos críticos, o desenvolvimento de um modelo preditivo que estime os tempos de espera e que caracterize pacientes é extremamente importante, pois oferece soluções eficazes para melhorar a qualidade e eficiência dos cuidados oferecidos.

2.1 Viabilidade

Este TFC responde diretamente a uma necessidade real identificada no CMRA: otimizar a capacidade de gestão de internamentos e tratamentos e a caracterização de pacientes de modo a melhorar o atendimento.

Este TFC tem fundamento:

- na cooperação entre o CMRA e o DEISI que resultou na conclusão de diversos TFC nos últimos dois anos letivos;
- num protocolo de cooperação científica entre o Centro de Medicina de Reabilitação de Alcoitão e a COFAC - Cooperativa de Formação e Animação Cultural, CRL. que será firmado em breve;
- em dados fornecidos pela Diretora da Unidade de Gestão de Utentes do CMRA, Engenheira Mariana Matos, assim como a sua contínua colaboração com a equipa deste TFC, ao longo deste ano lectivo;
- no reforço da adequação do modelo a desenvolver por este TFC, aos desafios específicos do centro em resultado do diálogo com a instituição, de modo a garantir que a solução a obter responda às prioridades do CMRA.

Além do potencial suporte na previsão dos tempos de espera e na caracterização de pacientes, o modelo a desenvolver por este TFC representa também uma alternativa financeiramente viável, proporcionando resultados precisos sem custos. A implementação de um modelo preditivo poderá eliminar a necessidade de soluções mais complexas e onerosas, como a implementação de sistemas adicionais ou o aumento das equipas, permitindo que o centro faça uma gestão mais otimizada dos seus recursos.

Adicionalmente, é importante observar que a flexibilidade do modelo a obter, embora desenvolvido com foco nas necessidades do CMRA, permitirá uma fácil adaptação a diferentes conjuntos de dados.

2.2 Pertinência

Este trabalho irá resultar na implementação de um modelo que permitirá trazer melhorias diretas ao planeamento de internamento de pacientes do CMRA e à caracterização dos pacientes, com potencial de aumentar a eficiência e a qualidade dos serviços oferecidos.

- **A previsão de tempos de espera de doentes para internamento e tratamento** vai permitir ao CMRA otimizar o atendimento de doentes, permitindo uma distribuição mais eficiente dos recursos e, conseqüentemente, reduzindo o tempo de espera para os pacientes.
- **A caracterização de pacientes**, permite a otimização da distribuição de recursos e permitirá um atendimento mais organizado.
- **Com melhor planeamento**, o centro poderá atender a um maior número de pacientes sem comprometer a qualidade dos cuidados prestados.

Assim sendo, a implementação do modelo preditivo neste TFC irá abordar uma necessidade do CMRA, contribuindo para a sustentabilidade e a melhoria contínua dos serviços de saúde prestados.

Outro aspeto importante resulta da relevância do CMRA no contexto da investigação científica na área da reabilitação. Esta relevância do CMRA resulta do número de doentes que atende, comparativamente com outras unidades de saúde na Europa. O modelo preditivo a aplicar por este TFC permitirá, mais tarde, que se desenvolva investigação no que toca ao seu desenvolvimento, conseqüências, melhorias a implementar, possível implementação a outros processos clínicos ou unidades hospitalares.

3 - Conceitos Fundamentais

O nosso objetivo, neste ponto, é apresentar fundamentos teóricos necessários para uma melhor compreensão do desenvolvimento deste trabalho, incluindo a discussão sobre o conceito de Machine Learning e a exploração de técnicas e modelos relevantes que poderão vir a ser utilizados ao longo do projeto.

3.1 Conceitos Teóricos

Nesta secção, queremos apresentar, e analisar, as técnicas que irão ser aplicadas nos capítulos seguintes, com principal destaque para os princípios de Machine Learning e a importância da preparação dos dados. Estes conceitos são essenciais para garantir a eficácia do modelo preditivo que queremos desenvolver, assim como, compreender as métricas que poderão avaliar o seu desempenho.

3.1.1 Machine Learning

Machine Learning é uma área da Inteligência Artificial (IA) que desenvolve algoritmos capazes de aprender, tomar decisões e fazer previsões ou classificações baseadas em dados.

A eficácia destes modelos depende dos dados a utilizar, quanto mais relevantes, diversos e de qualidade estes forem, melhor será o desempenho do modelo. Para garantir a eficácia, é necessário um grande volume de dados, permitindo ao modelo detetar padrões e relações que podem ser difíceis de perceber manualmente.

3.1.2 Machine Learning na saúde

Na área da saúde, o principal objetivo do Machine Learning é possibilitar o processamento eficiente de grandes quantidades de dados, a identificação de padrões e a obtenção de informações essenciais para os profissionais de saúde.

Em contextos hospitalares, por exemplo, a gestão de tempo e recursos é essencial, fazendo com que as técnicas sejam determinantes, para otimizar a distribuição de profissionais, prever a necessidade de internamento e classificar os pacientes de acordo com as suas patologias.

Ao longo do desenvolvimento do projeto, este ponto será revisto e, para que seja mais perceptível, na secção 4.1 estão apresentados artigos que ajudam a entender melhor o papel da aplicação de ML na área da saúde.

3.1.3 Tipos de Machine Learning

O Machine Learning está dividido em dois tipos principais de aprendizagem. Estes são essenciais para entender como os modelos são desenvolvidos e treinados:

- **Aprendizagem Supervisionada** – Neste tipo de aprendizagem, o modelo é treinado a partir de dados rotulados, ou seja, as variáveis independentes (características) e as dependentes (rótulos), são previamente conhecidas. O objetivo é criar um modelo capaz de fazer previsões precisas, com base em novos dados. Esta aprendizagem, é especialmente útil, em problemas de classificação e regressão.
- **Aprendizagem Não Supervisionada** – Trata-se de um tipo de aprendizagem usada quando os dados não estão rotulados. O objetivo é identificar padrões ocultos nos dados, sem a necessidade de rótulos específicos.[5]

3.1.4 Preparação dos Dados

Antes dos modelos de Machine Learning serem utilizados, é necessário preparar os dados. A sua qualidade tem um papel essencial no desempenho dos modelos preditivos. Assim sendo, dados inconsistentes, incompletos ou redundantes, podem comprometer o desempenho de qualquer algoritmo.

Esta preparação, é um passo muito importante no desenvolvimento de modelos. Os principais processos incluem:

- **Limpeza dos dados** – remoção de valores ausentes, duplicados ou incorretos, de modo a garantir a consistência.
- **Transformação de dados** – normalização ou padronização para que as variáveis estejam todas na mesma escala.
- **Feature engineering** – seleção e criação de variáveis com maior impacto no modelo.
- **Divisão de Dados** – separação em conjuntos de treino e teste, de forma a evitar o overfitting (quando o modelo decora os dados de treino e não funciona bem para novos dados).

3.1.5 Métricas de Desempenho

Para que seja possível classificar o desempenho do modelo é necessário utilizar certas métricas para avaliar a eficácia do mesmo.

As principais métricas que poderão vir a ser utilizadas ao longo do desenvolvimento do trabalho são:

- **Coeficiente de determinação (R^2)** - Medida estatística que mostra o grau de proximidade entre os dados e os valores ajustados pelo modelo de regressão. Varia

entre 0 e 1, sendo que, quanto mais próximo de 1 o valor estiver, melhor é o ajuste e menor o erro. [6]

$$R^2 = 1 - \frac{\text{Soma dos erros quadrados}}{\text{Soma total dos quadrados}} \quad (1)$$

- **Erro Quadrático Médio (MSE)** - Representa a média das diferenças quadráticas entre os valores reais e os valores previstos pelo modelo. Quanto maior for o resultado, pior é o desempenho do modelo, já que este indica que as previsões estão longe dos valores reais. [6]

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (2)$$

- **Raiz do Erro Quadrático Médio (RMSE)** - É a raiz quadrada do erro quadrático médio e mede o desvio padrão das diferenças entre os valores reais e os valores previstos. Quanto menor for o resultado, melhor é o desempenho do modelo. [6]

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3)$$

- **Erro Médio Absoluto (MAE)** - Esta métrica calcula a média das diferenças absolutas entre os valores reais e os valores previstos. [7]

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (4)$$

- **Erro Médio (ME)** - É uma métrica que calcula apenas a média das diferenças entre os valores reais e os valores previstos pelo modelo, dando assim uma visão mais geral do seu desempenho. Quanto mais próximo de 0 o seu resultado estiver, melhor é o desempenho do modelo. [8]

$$ME = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (5)$$

- **Erro Percentual Absoluto Médio (MAPE)** - Mede a precisão do modelo ao calcular a média dos erros percentuais absolutos. Este calcula a soma dos erros de previsão absolutos, expressos em percentagem, e divide pelo número total de valores reais. Quanto mais próximo o resultado estiver de 0 melhor é o desempenho do modelo. [6]

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (6)$$

- **Soma de Erro Absoluto (SAE)** - É uma medida que calcula a soma das diferenças entre os valores reais e os valores previstos pelo modelo, ajudando a compreender os erros acumulados num conjunto de dados. [9]

$$\sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

A utilização destas métricas, oferece uma visão mais concreta do desempenho do modelo, permitindo ajustar e comparar diferentes abordagens durante o desenvolvimento do trabalho.

As métricas descritas serão consideradas durante a avaliação dos modelos a implementar neste TFC. No entanto, a escolha de métricas a utilizar, dependerá igualmente dos dados e do modelo a obter.

3.1.6 Modelos a utilizar

De modo a facilitar a compreensão dos modelos que poderão ser utilizados ao longo do projeto, seguem-se observações sobre os mesmos. Estes modelos, representam diversas maneiras de abordar os problemas propostos no âmbito do trabalho: previsão de tempos de espera e caracterização dos pacientes.

Depois de o CMRA nos disponibilizar o conjunto de dados com que vamos trabalhar, faremos uma análise detalhada desses dados. Essa análise permitirá adequar o modelo preditivo a desenvolver à abordagem mais adequada.

Regressão Linear

A regressão linear, é um modelo estatístico utilizado para analisar a existência de uma relação linear entre variáveis e prever valores. Este modelo baseia-se na relação entre uma variável dependente (variável a prever) e uma ou mais variáveis independentes (utilizadas para a previsão). O seu objetivo principal, é identificar as relações entre as variáveis, ajustando uma linha reta que representa os dados. [10]

A equação da regressão linear simples é:

$$y = \beta_0 + \beta_1 \cdot x + \epsilon \quad (8)$$

Onde:

- x é a variável independente (utilizada para prever y);
- y é a variável dependente (variável a prever);
- β_0 é a ordenada na origem (valor de y quando $x=0$);
- β_1 é o coeficiente que mede o impacto da variável x sobre a y ;

- ϵ é o erro (diferença entre valores reais e previstos).

Tipos de regressão linear:

- **Regressão Linear Simples** – apenas uma variável independente é considerada para a previsão. (Figura 2)
- **Regressão Linear Múltipla** – são consideradas várias variáveis independentes para a previsão. (Figura 3)



Figure 2: Regressão Linear Simples

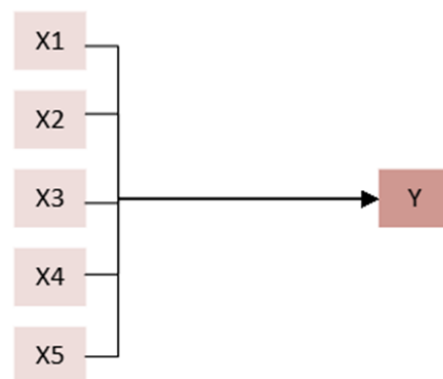


Figure 3: Regressão Linear Múltipla

Árvores de Decisão

Uma árvore de decisão é representada por uma estrutura semelhante a um fluxograma, onde cada ramificação interna representa uma característica e cada extremidade representa o valor previsto, permitindo visualizar de forma simples como o modelo chega a determinada conclusão.

Esta é uma abordagem preditiva e é utilizada para problemas de classificação e regressão. As árvores de decisão são construídas através de uma abordagem algorítmica que identifica maneiras sobre como dividir um conjunto de dados dependendo das condições.

Existem diferentes tipos de árvores de decisão, dependendo do que queremos obter:

- **Árvores de classificação** - onde a variável alvo pode assumir um conjunto discreto de valores.
- **Árvores de regressão** - onde a variável alvo pode assumir valores contínuos.

Estas árvores utilizam algoritmos que identificam o melhor atributo para dividir os dados em cada etapa, otimizando a precisão das previsões.[11]

De seguida, na figura 5, está apresentado um exemplo de uma árvore de decisão:

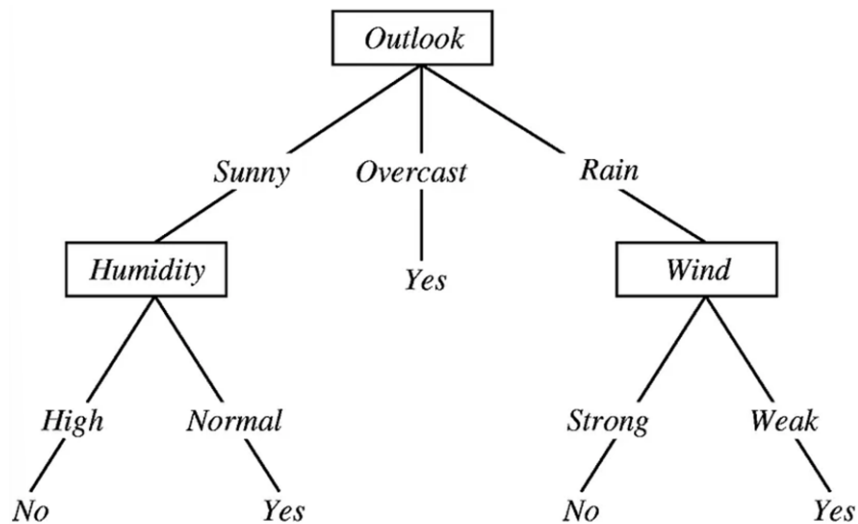


Figure 4: Exemplo de uma árvore de decisão

Random Forest

Random Forest é um algoritmo utilizado para classificação e regressão, que funciona através da criação de várias árvores de decisão durante o treino do modelo. Cada árvore é construída com um subconjunto do conjunto inicial de dados, que é selecionado aleatoriamente.

No caso da classificação, a previsão final é selecionada pela classe que mais aparece nas árvores, e no caso da regressão, pela média das previsões.

É um modelo que ajuda a reduzir o overfitting e melhora a precisão preditiva. Ao contrário das árvores de decisão, que avaliam todas as possíveis divisões de características, este seleciona aleatoriamente um subconjunto de características para cada divisão, resultando num melhor desempenho. A aleatoriedade na seleção do subconjunto de características é essencial para reduzir a correlação entre as árvores e garantir um modelo mais fiável. [12]

Na imagem a seguir, figura 6, pode-se observar o funcionamento do Random Forest.

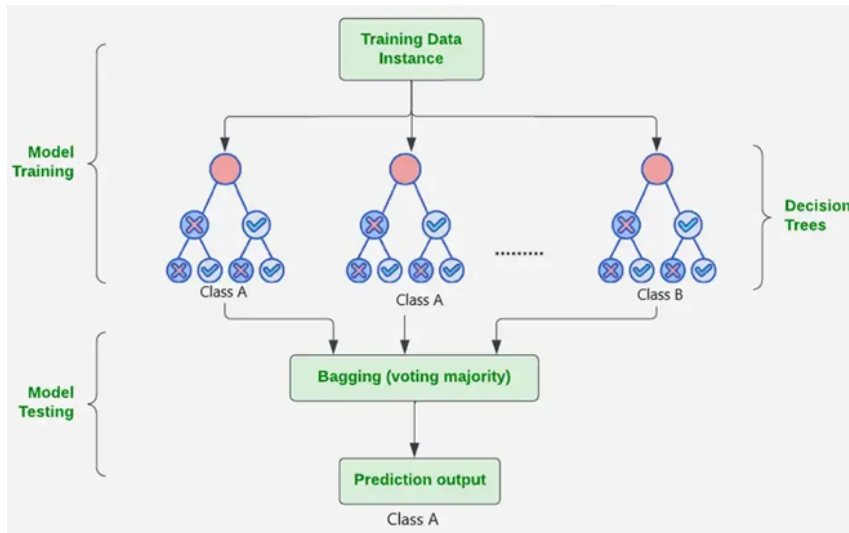


Figure 5: Demonstração do funcionamento do algoritmo Random Forest

Tal como mostra a imagem, o treino do modelo envolve a criação de várias árvores de decisão com dados selecionados aleatoriamente. Cada árvore faz uma previsão e, durante o teste, as previsões são combinadas através do processo de bagging, onde a classe mais frequente é escolhida para previsão total.

Support Vector Machine (SVM)

O SVM (Support Vector Machine) é um algoritmo particularmente eficaz quando usado em datasets mais pequenos, pois este pode ser demorado no processamento de datasets grandes.

O seu principal objetivo é encontrar o melhor hiperplano que separa as classes de dados num espaço N-dimensional, maximizando a distância entre elas. Esse hiperplano é o limite de decisão que determina a classe de um novo ponto.

Quando os dados não podem ser separados de forma linear, utiliza-se o kernel (função matemática que calcula a semelhança entre pontos, facilitando a separação entre classes) para transformar os dados num espaço de maior dimensão. O modelo pode ser utilizado tanto em problemas de classificação como de regressão. No entanto, tende a ser mais eficaz em problemas de classificação. [13]

Na imagem a seguir, figura 7, pode-se observar uma aplicação de SVM. Os pontos azuis e verdes representam as classes de dados e o hiperplano (linha vermelha) é o que separa as classes. Através da imagem torna-se mais fácil entender como este modelo trabalha para encontrar o melhor limite.

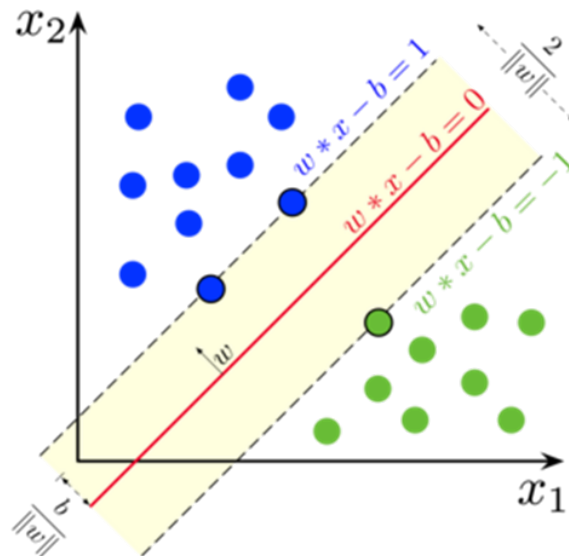


Figure 6: Exemplo de uma aplicação de SVM

K- Nearest Neighbours (KNN)

O KNN é um algoritmo muito usado para problemas de classificação e de previsão, baseando-se no princípio que, os pontos semelhantes estão próximos uns dos outros. Na classificação, o rótulo de um ponto é atribuído com base no rótulo mais frequente entre os k vizinhos mais próximos, enquanto na regressão, utiliza-se a média dos pontos vizinhos como previsão. Para medir as distâncias e identificar os pontos mais próximos, a métrica mais utilizada é a distância euclidiana (distância mais curta entre dois pontos).

Como o modelo armazena apenas os dados de treino, exige muita memória e pode não ser eficiente em grandes conjuntos de dados. Assim sendo, a escolha do k é muito importante, valores pequenos podem levar a overfitting e valores grandes aumentam o bias. O melhor ajuste é, geralmente, alcançado através da validação cruzada.[14]

Na imagem abaixo, figura 8, é possível observar o processo do algoritmo:

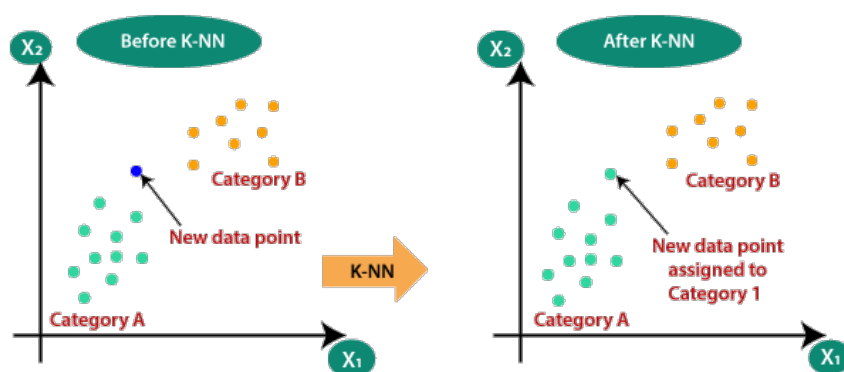


Figure 7: Aplicação do algoritmo KNN

O ponto a identificar (representado a azul) é classificado com base nos seus vizinhos

mais próximos. Antes da aplicação KNN o ponto não tem classe, e após a aplicação, é classificado, de acordo com os seus vizinhos mais próximos, na categoria A (representado a verde).

K-Means

O K-Means é um algoritmo destinado a resolver problemas de clustering, uma técnica que tem como objetivo agrupar pontos de dados semelhantes, tendo em conta as suas características. O clustering divide um conjunto de dados em clusters distintos, onde cada cluster consiste em pontos de dados mais semelhantes entre si.

Este é um algoritmo de agrupamento, utilizado para classificar dados não rotulados em grupos (clusters) tendo em conta a sua similaridade. O processo envolve a definição do valor ideal de K pontos centrais (centróides) através de um processo iterativo e, de seguida, atribui cada ponto ao seu centróide mais próximos criando assim os clusters. [15]

Na figura abaixo, figura 9, é possível observar o resultado a obter através do processo:

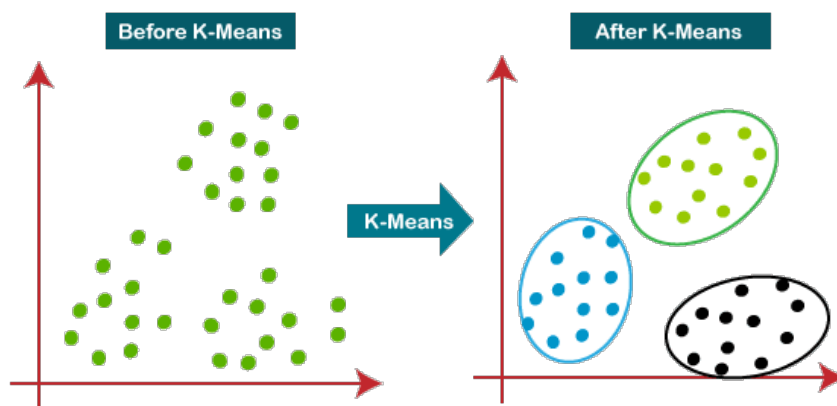


Figure 8: Resultado da aplicação do algoritmo K-Means

Este algoritmo será utilizado na caracterização dos pacientes, segmentando-os tendo em conta características semelhantes. O objetivo é agrupar os pacientes de maneira que cada grupo represente um conjunto distinto de pacientes, facilitando a gestão de recursos e permitindo ao CMRA tomar decisões mais informadas e adequadas para cada tipo de característica.

4 - Estado da Arte

Este capítulo tem como objetivo analisar as vantagens e desafios da implementação de modelos de Machine Learning (ML) na previsão de tempos de espera e na caracterização de pacientes no CMRA, com base numa pesquisa sobre modelos já desenvolvidos para objetivos semelhantes. A partir da pesquisa realizada, foram identificados diversos modelos e abordagens que utilizam técnicas de ML, variando em complexidade e eficácia. Estes modelos refletem diferentes estratégias para otimizar a distribuição de recursos e aumentar a eficiência nos serviços de saúde.

Nos últimos anos, a aplicação de algoritmos de ML na saúde, tem gerado resultados promissores, principalmente devido ao aumento da disponibilidade de dados clínicos e à capacidade dos modelos de lidar com a complexidade e variabilidade dos dados. Esta evolução permitiu o desenvolvimento de diversos métodos capazes de otimizar a gestão de recursos hospitalares e melhorar o atendimento aos pacientes.

Diversos estudos analisam diferentes processos de Machine Learning como, por exemplo, Regressão Linear e K-Nearest Neighbors (KNN), Gradient Boosting Decision Trees (GBDT) e Redes Neurais (NN). Cada uma destas abordagens, apresenta vantagens específicas dependendo do tipo de dados, número de variáveis envolvidas e nível de precisão exigida. A seguir, será realizada uma revisão detalhada dos estudos mais relevantes encontrados sobre o tema, com o objetivo de fornecer uma visão aprofundada do estado de arte na previsão de tempos de espera e na classificação de pacientes no contexto hospitalar.

4.1 Artigos Analisados

Esta é uma lista contendo os artigos que analisei e que exploram a aplicação de Machine Learning (ML) em contextos semelhantes ao do CMRA, com foco na previsão de tempos de espera e classificação de pacientes.

Artigo 1 – Anat Ratnovskya e outros [16]

"Statistical learning methodologies and admission prediction in an emergency department", investiga aplicação de Machine Learning (ML) para prever a admissão hospitalar de pacientes no setor de emergências. O artigo, propõe um modelo preditivo baseado em dados iniciais de triagem. Para identificar os principais indicadores de desempenho (KPIs) que influenciam a previsão, são realizadas análises exploratórias (AED) e confirmatórias (ACD) dos dados. Os algoritmos aplicados incluem a Regressão Logística, a Árvore de Decisão e as Redes Neurais Artificiais (RNA). A métrica AUC (Área Sob a Curva) da curva ROC (Receiver Operating Characteristic), que avalia a taxa de verdadeiros e falsos positivos, é usada para medir a eficácia dos modelos, tendo o modelo de RNA, demonstrado o melhor desempenho preditivo.

Artigo 2 - Jeffin Joseph e outros [17]

"Machine Learning for Prediction of Wait Times in Outpatient Clinics", explora o uso

dos algoritmos Random Forest e Extreme Gradient Boosting (XGBoost) para prever tempos de espera e o tempo de processamento em clínicas em regime ambulatorio. As métricas utilizadas para avaliar o desempenho dos modelos incluem precisão, acurácia e a curva ROC. Para lidar com o desequilíbrio de classes, foi aplicada a técnica de super amostragem de vizinhança sintética (SMOTE), resultando numa melhoria significativa dos resultados. O XGBoost, especialmente quando combinado com SMOTE, obteve uma precisão de 0,86 e uma especificidade muito alta, destacando-se como uma ferramenta eficaz na gestão de filas e na melhoria da eficiência dos serviços de saúde.

Artigo 3 - Anton Pak e outros [18]

No artigo "Predicting waiting time to treatment for emergency department patients", são propostos algoritmos de ML para prever, com mais precisão, os tempos de espera em serviços de urgência. O estudo compara diferentes modelos, como LASSO (Least Absolute Shrinkage and Selection Operator), Random Forest e Quantile Regression, que superaram métodos tradicionais, como a média móvel, ao capturar, de forma mais eficaz, a variabilidade nos tempos de espera. Os resultados indicam que modelos como LASSO e Random Forest, são mais eficazes em cenários de alta procura, oferecendo previsões mais precisas e uteis.

Artigo 4 - A/Prof Katie Walker e outros [19]

O artigo "Emergency medicine patient wait time multivariable prediction models", apresenta a validação de modelos de ML para prever os tempos de espera em serviços de urgência, através dos dados de 12 hospitais na Austrália. A pesquisa compara modelos como Random Forest e Regressão Linear com modelos tradicionais, como a media móvel. Os modelos desenvolvidos para cada hospital, mostraram maior precisão, principalmente, em situações excepcionais como a pandemia de COVID-19. O estudo destaca a importância da validação externa dos modelos antes da sua implementação noutros hospitais, devido à limitação da transferência de modelos específicos entre locais.

Artigo 5 – Amjed Al-Mousa e outros [20]

O estudo "A machine learning-based approach for wait-time estimation in healthcare facilities with multi-stage queues", propõe uma abordagem baseada em ML, para estimar tempos de espera em hospitais, tendo em consideração, a complexidade das filas multifacetadas. A pesquisa utiliza algoritmos como Random Forest, Ridge Regression, LASSO, Elastic Net e Redes Neurais Artificiais (RNA). O modelo Random Forest obteve o menor valor de erro quadrático médio (RMSE), superando métodos tradicionais. O estudo destaca a capacidade dos modelos de ML de melhorar significativamente a precisão das previsões, principalmente, em cenários com múltiplas etapas de filas.

Artigo 6 - Parthasarathi Pattnayak e outros [21]

Este estudo, "Deep Learning based Patient Queue Time Forecasting in the Emergency Room", propõe um modelo baseado em deep learning (DL) para prever, com maior precisão, os tempos de espera em serviços de emergência, focando em pacientes de baixa gravidade. Motivado pela alta taxa de erro de modelos anteriores, o artigo utiliza técnicas de DL e teoria das filas de espera, para estimar os tempos de espera com base em dados de Registos Eletrônicos de Saúde (EHR). O modelo melhora a utilização de variáveis preditoras como hora e dia de chegada, e tempo de permanência na fila, por meio de otimizações como Adagrad, Adam, RMSprop e SGD. O modelo de DL mostrou-se mais eficaz que métodos convencionais, apresentando um erro menor e contribuindo

para a priorização dos pacientes.

Artigo 7 – Xiaoqing Li e outros [22]

No estudo "Prediction of outpatient waiting time", são explorados diferentes modelos de ML, incluindo Regressão Linear, Random Forest, Gradiente Boosting Decision Tree (GBDT) e K-Nearest Neighbors (KNN), para prever os tempos de espera num hospital pediátrico chinês. O Random Forest, obteve um desempenho superior no Departamento de Medicina Interna I, com um erro médio absoluto (MAE) 47,6% inferior ao da Regressão Linear. Já o GBDT, apresentou os melhores resultados em outros departamentos. O estudo enfatiza ainda a importância da engenharia de características e da validação cruzada de modo a melhorar a precisão dos modelos.

Artigo 8 - Suebsarn Ruksakulpiwat e outros [23]

"Machine learning-based patient classification system for adults with stroke: A systematic review", explora o uso de técnicas de ML na classificação de pacientes adultos com AVC. São discutidas abordagens supervisionadas, não supervisionadas e por reforço, com destaque para o uso de DL, que obteve os melhores resultados na previsão de tempo de detecção e de internamento de pacientes com AVC.

Artigo 9 - Shatha Melhem e outros [24]

O estudo "Patient care classification using machine learning techniques", explora a classificação de pacientes (internados ou em ambulatório), utilizando técnicas de ML com os dados EHR de um hospital indonésio. Foram implementados modelos como SVM, Árvore de Decisão, Random Forest e KNN, destacando o Random Forest como o mais eficaz, com uma precisão de 77%. O estudo mostra que a classificação de pacientes, pode melhorar a eficiência dos serviços de saúde, reduzindo tempos de espera e erros nas decisões.

De modo a que seja mais fácil a comparação dos artigos segue-se uma tabela comparativa (Tabela 2):

Título do Artigo	Ano	Objetivo	Modelos e Desempenho	Dados	Utiliza- dos
"Statistical learning methodologies and admission prediction in an emergency department"	2021	Prever admissão hospitalar em emergência	Regressão Logística, Árvore de Decisão, RNA (melhor desempenho)	Dados de triagem inicial	
"Machine Learning for Prediction of Wait Times in Outpatient Clinic"	2022	Prever tempo de espera e tempo de processamento em clínicas ambulatoriais	Random Forest, XGBoost (um melhor desempenho com SMOTE)	Dados de consultas ambulatoriais	
"Predicting waiting time to treatment for emergency department patients"	2021	Prever tempo de espera em serviços de emergência	Quantile Regression, LASSO (melhores previsões)	Dados de um serviço de emergência	
"Emergency medicine patient wait time multivariable prediction models: a multicentre derivation and validation study"	2021	Prever tempos de espera em múltiplos serviços de emergência	Random Forest, Regressão Linear (melhor desempenho)	Dados de serviços de emergência na Austrália	12 de
"A machine learning-based approach for wait-time estimation in healthcare facilities with multi-stage queues"	2023	Prever tempos de espera em estabelecimentos de saúde com filas multifacetadas	Random Forest (melhor desempenho com menor RMSE)	Dados de transações de um grande hospital	
"Deep Learning based Patient Queue Time Forecasting in the Emergency Room"	2023	Prever com precisão o tempo de espera dos pacientes em salas de emergência	Deep Learning (superou as técnicas convencionais)	Dados de registros eletrônicos de saúde (EHR)	
"Prediction of outpatient waiting time: using machine learning in a tertiary children's hospital"	2023	Prever o tempo de espera num hospital pediátrico	GBDT (melhor desempenho), Random Forest, KNN	Dados do sistema de informação hospitalar (HIS)	
"Machine learning-based patient classification system for adults with stroke: A systematic review"	2023	Explorar o uso de ML na classificação de doentes com AVC	Deep Learning (melhores resultados na detecção precoce e internamento de pacientes com AVC, no entanto não houve consenso sobre o algoritmo mais eficaz)	Dados clínicos de pacientes com AVC (detalhes não especificados)	
"Patient care classification using machine learning techniques"	2021	Classificar pacientes (internados ou ambulatoriais) para melhorar a gestão hospitalar	Random Forest (alcançou 77% de precisão), SVM, Árvore de Decisão, KNN	Dados EHR (resultados dos testes laboratoriais dos doentes) de um hospital indonésio	

Table 1: Tabela de Comparação dos Estudos sobre Previsão de Tempos de Espera e Classificação de Pacientes

Este capítulo analisou a aplicação de modelos de Machine Learning (ML) na previsão de tempos de espera e na classificação de pacientes, com base em estudos relevantes. A pesquisa revelou que algoritmos como Random Forest, Redes Neurais Artificiais (RNA), Regressão Linear, XGBoost, Deep Learning e GBDT, apresentaram resultados positivos, dependendo do tipo de dados e do contexto em que foram aplicados.

Modelos como o Random Forest, que se destacou em diversos estudos pela sua eficácia na previsão de tempos de espera e classificação de pacientes, mostram ser uma boa escolha no desenvolvimento do modelo proposto.

Resumindo, os modelos de ML analisados, oferecem soluções favoráveis na otimização da gestão de recursos hospitalares e na melhoria da eficiência no atendimento aos pacientes.

5 - Solução Proposta

A solução proposta para a realização deste projeto, consiste na aplicação e implementação de um modelo preditivo, utilizando técnicas de Machine Learning. O objetivo é prever os tempos de espera e caracterizar os pacientes de modo a comparar as métricas de cada um dos modelos estudados e escolher a melhor solução.

Apesar da importância da aplicação de Machine Learning, o sucesso do modelo não dependerá apenas disso. O processo de preparação dos dados será o primeiro passo a ser dado, após a obtenção dos dados, e incluirá as seguintes etapas:

- **Limpeza**- remoção de valores ausentes, duplicados ou inconsistentes.
- **Transformação** – normalização dos dados para garantir a sua coerência.
- **Tratamento de informações** – aplicação de técnicas de pré-processamento de modo a garantir que os dados estão no formato adequado para a construção do modelo.

O próximo passo será a realização da análise exploratória dos dados. Nesta fase, a distribuição e o comportamento das variáveis, serão estudados através de visualizações gráficas e análises estatísticas, permitindo a identificação de padrões, a detecção de outliers e uma melhor compreensão do comportamento das variáveis. Esta análise é essencial para a detecção de relações entre variáveis e para um melhor entendimento do conjunto de dados. Através da análise de correlação entre as variáveis, será possível selecionar as mais importantes para o modelo.

Após a análise exploratória dos dados, será realizado o processo de feature engineering. Este consiste na transformação e seleção de variáveis para criar um modelo preditivo, com o objetivo de descobrir padrões ocultos nos dados e melhorar o desempenho do modelo. Nesta fase, serão aplicadas técnicas, como a normalização ou padronização de dados, entre outras, para que seja possível ajustar as escalas e destacar as variáveis mais relevantes para o modelo.

Com os dados já preparados, o próximo passo será a construção e o treino do modelo. Nesta fase, serão escolhidos e treinados os algoritmos mais adequados à resolução do problema, e os dados serão divididos, aleatoriamente, em conjuntos de treino e teste, de modo a garantir uma avaliação eficaz do modelo, evitando problemas como o overfitting e underfitting. A divisão dos dados será feita da seguinte forma:

- **70% dos dados** serão destinados ao conjunto de treino, permitindo ao modelo aprender os padrões presentes nas variáveis.
- Os restantes **30% dos dados** serão destinados ao conjunto de teste, utilizado para avaliar a capacidade do modelo perante novos dados.

De seguida, serão aplicados alguns dos modelos referidos nos conceitos teóricos, e será feita uma comparação dos seus desempenhos de modo a escolher o mais adequado. A avaliação do modelo escolhido será realizada, através da análise do seu

desempenho utilizando métricas, como o coeficiente de determinação, o erro quadrático médio, entre outras, de modo a assegurar a capacidade de previsão do modelo.

Por fim, com o modelo validado, a sua implementação poderá ocorrer, assegurando que este está preparado para ser utilizado de forma eficaz.

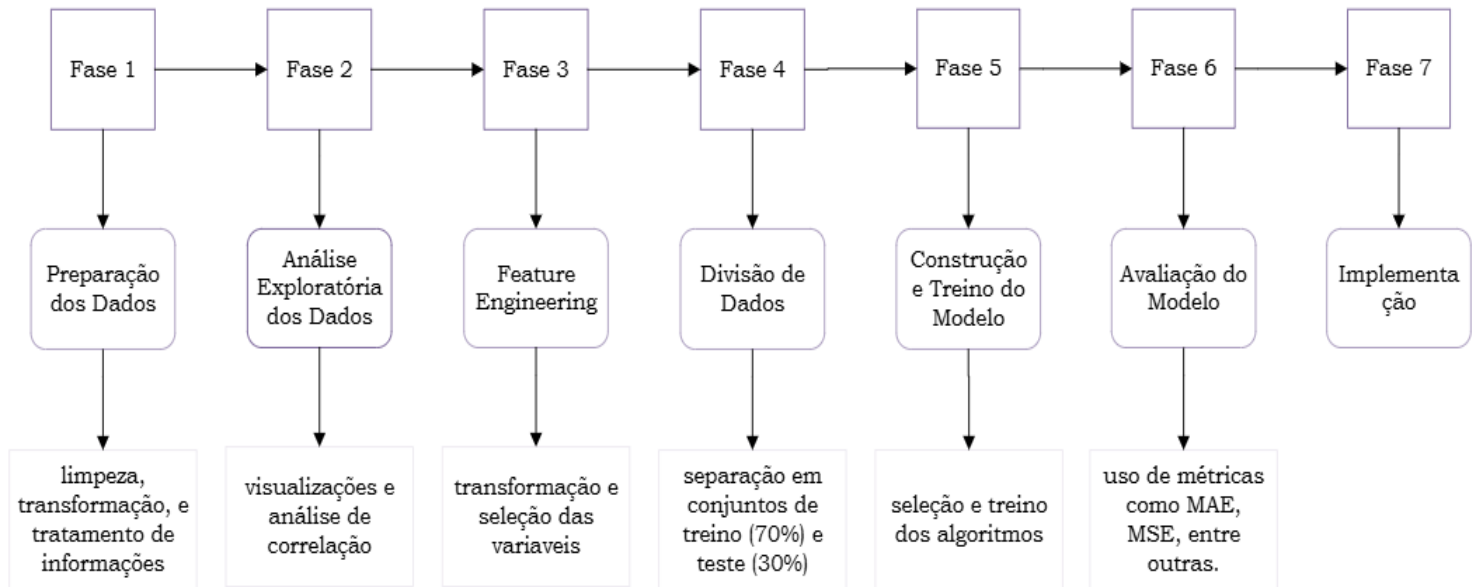


Figure 9: Fases propostas para a resolução do problema

5.1 Dados a Utilizar

A preparação deste TFC iniciou-se no início de setembro de 2024, com o convite da Engenheira Mariana Mota à Professora Iolanda Velho, para esta colaboração. Na altura antecipávamos que o CMRA partilharia, de uma forma célere e simples, o conjunto de dados na base do desenvolvimento do modelo preditivo. No entanto, colocou-se entretanto a necessidade de obter a aprovação da comissão de Ética do CMRA, que só se reúne no dia 4 de dezembro de 2024.

Assim, com base na análise feita a um conjunto similar de dados, proveniente do CMRA, e no contexto de um problema científico similar ao nosso, apresentamos variáveis que julgamos serem relevantes:

- ID_UT – Id do utente
- Sexo – Sexo do utente
- Data de Nascimento – Data de nascimento do utente
- Conc Residência – Concelho de residência do utente
- Data Início – Data do início do internamento
- Data Fim – Data de término do internamento

- TDI – Tempo de internamento (diferença entre Data fim e Data Início)
- Diagnóstico – Patologia identificada no utente
- DestinoPosAlta – Destino do paciente após alta médica

Ao utilizar variáveis detalhadas do historial de saúde dos utentes, o modelo poderá identificar padrões e relações entre estas. Dessa forma, poderá ser possível antecipar não apenas a necessidade imediata de internamento, como também estimar a possível evolução dos casos, prevendo, por exemplo, a duração esperada de recuperação e o tipo de cuidados necessários após a alta.

Este tipo de análise preditiva permite aos profissionais de saúde uma visão mais abrangente sobre cada caso. Através da análise do percurso dos pacientes, o sistema poderá prever tempos de espera, por exemplo, indicando quanto tempo um paciente terá de esperar para ser atendido, o que se traduz em decisões mais fundamentadas e uma gestão hospitalar mais eficaz.

O projeto tem então como objetivo explorar essas variáveis de modo que seja possível prever os tempos de espera e caracterizar os pacientes. Esta abordagem permitirá, principalmente, otimizar recursos e melhorar a qualidade dos cuidados prestados.

5.2 Ferramentas a utilizar

Para o desenvolvimento deste trabalho será usada a linguagem Python pois, para além de ser a mais praticada ao longo da licenciatura, é também a mais utilizada em trabalhos direcionados a machine learning.

Quanto à escolha de ambiente de desenvolvimento Jupyter Notebook, esta justifica-se pela sua capacidade de ser utilizado sem acesso à internet e pela sua eficiência que permite a junção de texto com código, garantindo a organização do trabalho. Além disso, possibilita, ainda, a execução de cada célula de código individualmente, tornando o processo mais dinâmico ao permitir testar e ajustar o código de forma isolada, sem a necessidade de o executar todo de uma vez.

Para o processamento, manipulação, visualização dos dados e implementação do modelo, serão utilizadas as bibliotecas:

- **Numpy** – Utiliza arrays multidimensionais eficientes, realiza operações matemáticas sem necessidade da criação de loops detalhados, disponibiliza variadas funções algébricas, estatísticas, entre outras, e serve ainda como base para diversas bibliotecas de Machine Learning, como, por exemplo, o Pandas. [25]
- **Pandas** – Permite a manipulação e organização de grandes volumes de dados de forma mais eficiente, possibilita a leitura de ficheiros de diversos formatos (csv, excel, json, entre outros) e fornece ainda ferramentas que permitem lidar com diferentes tipos de dados, garantindo a sua consistência. [26]
- **Matplotlib** – Facilita a criação de diversos tipos de gráficos, oferecendo ferramentas para personalizar e ajustar elementos gráficos consoante o que é pretendido obter. [27]

- **Scikit-learn** – Oferece uma grande coleção de algoritmos (como regressão, classificação, entre outros) e, além de incluir várias métricas que possibilitam a avaliação do desempenho do modelo, disponibiliza, também, ferramentas para a normalização e seleção de características, otimizando o processo. [28]

Estas ferramentas e bibliotecas não otimizam apenas o desenvolvimento e a análise de dados, como também garantem a consistência dos resultados e a clareza no processo de documentação.

5.3 Abrangência

Neste projeto, diversas Unidades Curriculares serão essenciais, refletindo a importância das aprendizagens adquiridas, ao longo da licenciatura, para a realização do TFC. Destacam-se as Unidades Curriculares de Análise Exploratória de Dados, Aprendizagem Automatizada I e II, e Linguagens de Programação, fundamentais para o tratamento e estruturação dos dados, desenvolvimento de modelos preditivos e a implementação de algoritmos eficazes.

Na fase inicial, conhecimentos de Probabilidades e Estatística serão cruciais para a análise estatística, garantindo a precisão dos resultados obtidos. Além disso, na fase seguinte, serão aplicadas técnicas de modelação e otimização de modelos, com o suporte de métodos aprendidos em Inteligência Artificial, que permitirão explorar novas abordagens para a resolução do problema.

Desta forma, pode-se afirmar que todas as Unidades Curriculares, mesmo as não mencionadas, trarão uma componente importante para a realização do projeto e para a qualidade da solução a ser desenvolvida.

6 - Método e Planeamento

Durante o desenvolvimento do projeto, ao longo do ano letivo, será seguido um plano de trabalho, que servirá como referência para a realização do projeto.

Este plano está dividido em três fases principais, com os seguintes pontos de referência:

- Entrega do Relatório Intercalar do 1º semestre (01/12/2024)
- Entrega do Relatório Intercalar do 2º semestre (13/04/2025)
- Entrega do Relatório Final (27/06/2025)

Estas fases têm objetivos específicos a seguir, estes correspondem a:

- **1ª fase – Entrega do Relatório Intercalar do 1º semestre**

- Identificação do problema
- Leitura e pesquisa bibliográfica
- Desenvolvimento da proposta de solução
- Definição de conceitos fundamentais

- **2ª fase – Entrega do Relatório Intercalar do 2º semestre**

- Recolha e preparação dos dados
- Análise exploratória dos dados
- Análise e implementação dos modelos a utilizar

- **3ª fase – Entrega do relatório Final**

- Registo dos resultados obtidos
- Melhoria do Modelo
- Teste e validação do modelo
- Análise dos resultados obtidos
- Conclusão do relatório

Estas são as fases previstas. No entanto, ao longo do desenvolvimento do trabalho, será possível ajustar, consoante os desafios que forem surgindo, e detalhar melhor cada passo a ser executado.

Na figura 10, apresenta-se um diagrama de Gantt, que mostra o tempo previsto para o desenvolvimento de cada etapa. Este cronograma será atualizado ao longo do desenvolvimento do projeto, conforme necessário.

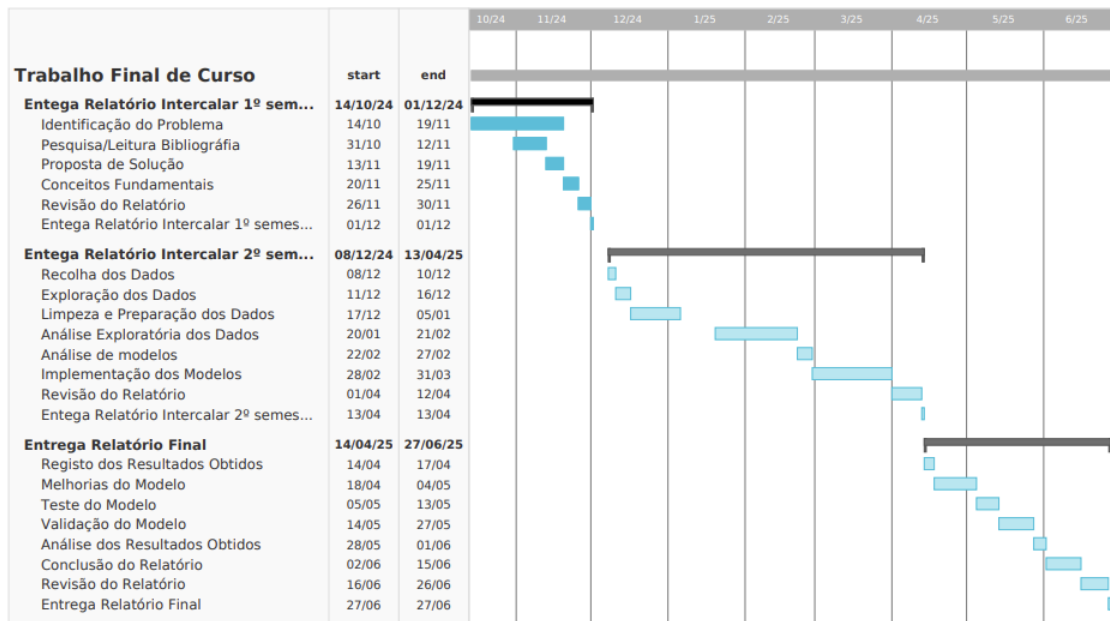


Figure 10: Diagrama Gantt - 1º Semestre

Bibliografia

- [1] João P. Matos-Carvalho. *The Lusófona L^AT_EX Template User's Manual*. Lusófona University. 2024. URL: <https://github.com/jpmcarvalho/UL-Thesis>.
- [2] Santa Casa da Misericórdia de Lisboa. *Centro de Medicina de Reabilitação de Alcoitão*. 2023. URL: <https://scml.pt/media/noticias/centro-de-medicina-e-reabilitacao-de-alcoitao/> (visited on 10/2024).
- [3] Centro de Reabilitação de Alcoitão. *Sobre nós - Centro de Reabilitação de Alcoitão*. URL: <https://cmra.scml.pt/sobre-nos/> (visited on 10/2024).
- [4] Centro de Reabilitação de Alcoitão. *Como fazemos - Centro de Reabilitação de Alcoitão*. URL: <https://cmra.scml.pt/como-fazemos/> (visited on 10/2024).
- [5] IBM. *What is machine learning?* URL: <https://www.ibm.com/topics/machine-learning>.
- [6] Felipe Azank. *Como avaliar seu modelo de regressão*. 2020. URL: <https://medium.com/turing-talks/como-avaliar-seu-modelo-de-regress%C3%A3o-c2c8d73dab96>.
- [7] Akshita Chugh. *MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better?* 2020. URL: <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e>.
- [8] Statistics How To. *Mean Error: Definition*. URL: <https://www.statisticshowto.com/mean-error/>.
- [9] Mohamad. *SAE - Sum of Absolute Errors*. 2016. URL: <https://support.numxl.com/hc/en-us/articles/214286506-SAE-Sum-of-Absolute-Errors>.
- [10] Saishruthi Swaminathan. *Linear Regression — Detailed View*. 2018. URL: <https://towardsdatascience.com/linear-regression-detailed-view-ea73175f6e86>.
- [11] Prince Yadav. *Decision Tree in Machine Learning*. 2018. URL: <https://towardsdatascience.com/decision-tree-in-machine-learning-e380942a4c96>.
- [12] Carolina Bento. *Random Forests Algorithm explained with a real-life example and some Python code*. 2021. URL: <https://towardsdatascience.com/random-forests-algorithm-explained-with-a-real-life-example-and-some-python-code-affbfa5a942c>.
- [13] Wikipedia. *Support vector machine*. URL: https://en.wikipedia.org/wiki/Support_vector_machine.
- [14] Onel Harrison. *Machine Learning Basics with the K-Nearest Neighbors Algorithm*. 2018. URL: <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>.
- [15] Dishant kharkar. *Algoritmo de agrupamento K-Means*. 2023. URL: <https://medium.com/@dishantkharkar9/k-means-clustering-algorithm-ce4fbcac8fb0>.
- [16] Anat Ratnovsky et al. "Statistical learning methodologies and admission prediction in an emergency department". In: *Australasian Emergency Care* 24.4 (2021), pp. 241–247. ISSN: 2588-994X. DOI: <https://doi.org/10.1016/j.auec.2020.11.004>. URL: <https://www.sciencedirect.com/science/article/pii/S2588994X20301147>.

- [17] Jeffin Joseph et al. "Machine Learning for Prediction of Wait Times in Outpatient Clinic". In: *Procedia Computer Science* 215 (2022). 4th International Conference on Innovative Data Communication Technology and Application, pp. 230–239. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2022.12.026>. URL: <https://www.sciencedirect.com/science/article/pii/S187705092202097X>.
- [18] Anton Pak, Brenda Gannon, and Andrew Staib. "Predicting waiting time to treatment for emergency department patients". In: *International Journal of Medical Informatics* 145 (2021), p. 104303. ISSN: 1386-5056. DOI: <https://doi.org/10.1016/j.ijmedinf.2020.104303>. URL: <https://www.sciencedirect.com/science/article/pii/S1386505620305219>.
- [19] Katie Walker et al. "Emergency medicine patient wait time multivariable prediction models: a multicentre derivation and validation study". In: *medRxiv* (2021). DOI: 10.1101/2021.03.19.21253921. eprint: <https://www.medrxiv.org/content/early/2021/03/24/2021.03.19.21253921.full.pdf>. URL: <https://www.medrxiv.org/content/early/2021/03/24/2021.03.19.21253921>.
- [20] Mohammad Al-Dweik Amjed Al Moussa Hamza Al-Zubaidi. "A machine learning-based approach for wait-time estimation in healthcare facilities with multi-stage queues". In: *IET WILEY* (2023). DOI: 10.1049/smc2.12079.
- [21] Parthasarathi Pattnayak et al. "Deep Learning Based Patient Queue Time Forecasting in the Emergency Room". In: (2023), pp. 541–545. DOI: 10.1109/ICSSAS57918.2023.10331755.
- [22] Xiaoqing Li et al. "Prediction of outpatient waiting time: using machine learning in a tertiary children's hospital". In: *Translational Pediatrics* 12.11 (2023). ISSN: 2224-4344. URL: <https://tp.amegroups.org/article/view/119333>.
- [23] Zhou W Benjasirisan C Phianhasin L Schiltz NK Brahmabhatt S Ruksakulpiwat S Thongking W. "Machine learning-based patient classification system for adults with stroke: A systematic review". In: *Chronic Illn* (2023). DOI: 10.1177/17423953211067435.
- [24] Shatha Melhem, Ahmad Al-Aiad, and Muhammad Al-Ayyad. "Patient care classification using machine learning techniques". In: (May 2021), pp. 57–62. DOI: 10.1109/ICICS52457.2021.9464582.
- [25] Daniel Amorim. *Python — Pandas (Essenciais)*. 2023. URL: <https://medium.com/@dev.daniel.amorim/python-pandas-24ab58577de5>.
- [26] Francesca Fuentes. *NumPy: The Fundamental Tool for Data Science in Python*. 2024. URL: <https://medium.com/@m.franfuentes/numpy-the-fundamental-tool-for-data-science-in-python-fa2b605a3bf9>.
- [27] Matplotlib. *Matplotlib: Visualization with Python*. URL: <https://matplotlib.org/>.
- [28] Hugo Habbema. *Introdução ao Scikit-Learn*. 2024. URL: <https://medium.com/@habbema/introdu%C3%A7%C3%A3o-ao-scikit-learn-f00b7201dbf7>.