



**CIÊNCIAS
EMPRESARIAIS**

ESCOLA SUPERIOR
POLITÉCNICO SETÚBAL

VITOR EMANUEL
DE CARVALHO
RODRIGUES DE
ALMEIDA

**A RETENÇÃO DE MILITARES DOS
QUADROS PERMANENTES NA
FORÇA AÉREA – OS DADOS COMO
FONTE DE INFORMAÇÃO E
PREVISÃO**

Relatório de Dissertação do Mestrado em Ciência
de Dados para Empresas

ORIENTADORAS

Professora Doutora Helena Alexandra Couceiro
Feio de Almeida Penalva

Professora Doutora Sandra Cristina Dias Nunes

dezembro, 2024

VITOR EMANUEL
DE CARVALHO
RODRIGUES DE
ALMEIDA

**A RETENÇÃO DE MILITARES DOS
QUADROS PERMANENTES NA
FORÇA AÉREA – OS DADOS COMO
FONTE DE INFORMAÇÃO E
PREVISÃO**

JÚRI

Presidente: Professor Doutor, Vítor Manuel Meneses
Barbosa, ESCE

Orientador: Professora Doutora, Helena Alexandra
Couceiro Feio de Almeida Penalva, ESCE

Vogal Arguente: Professora Doutora Sandra Maria
Simões de Oliveira, ESCE

dezembro, 2024

Agradecimentos

Em primeiro lugar, quero deixar o meu profundo agradecimento a todos os que contribuíram, direta e indiretamente, para a conclusão deste trabalho e consequente etapa académica.

Agradecer, em particular, às minhas orientadoras, Professora Doutora Helena Penalva e Professora Doutora Sandra Nunes, pela confiança, orientação, disponibilidade e dedicação na realização deste trabalho. Sem elas a qualidade global desta dissertação não seria, na minha opinião, tão elevada.

Agradeço ao Sr. Major-General João Pereira pela visão e pelo incentivo na escolha deste tema. Conseguiu ver mais longe e, por isso, ficarei eternamente grato.

Grato à Força Aérea, na qualidade do Capitão Ricardo Pedro (elemento de ligação), por ter acedido à minha pretensão e ter disponibilizado os dados para a realização deste trabalho. Espero que seja um contributo para a obtenção de conhecimento que ajude a nossa Força Aérea a prosperar.

Por fim, e às pessoas mais importantes, Célia e Alice. Os meus pilares e aquelas a quem privei mais do meu tempo em família. Obrigado pela paciência, compreensão e por tudo o que me dão. Esta conquista também é vossa. Amo-vos.

Resumo

Este estudo investiga a problemática das saídas voluntárias dos militares dos Quadros Permanentes (QP) da Força Aérea Portuguesa, desenvolvendo um modelo preditivo para identificar militares com maior probabilidade de pedir abate. Utilizando uma abordagem quantitativa, recorreu-se a métodos de *machine learning*, nomeadamente Regressão Logística, *Naive Bayes*, Análise Discriminante Linear, Análise Discriminante Quadrática e Árvores de Decisão. A base de dados foi composta por 2176 militares, divididos entre aqueles que solicitaram abate e os que permaneceram em serviço. A análise identificou variáveis sociodemográficas e estruturais significativas, como a idade, a especialidade, o número de louvores e a nota da ficha de avaliação. O modelo de Regressão Logística (RL) apresentou a melhor performance, permitindo calcular um índice de probabilidade de saída, com utilidade prática para a gestão de recursos humanos. Apesar de limitações, como a baixa sensibilidade e o uso de um conjunto limitado de variáveis, o estudo oferece contributos importantes para a retenção de talento na Força Aérea e sugere a inclusão de mais variáveis e abordagens avançadas em investigações futuras.

Abstract

This study investigates the issue of voluntary departures of permanent staff (QP) from the Portuguese Air Force, developing a predictive model to identify military personnel with a higher likelihood of requesting departure. Using a quantitative approach, were used machine methods learning, namely Logistic Regression, Naive Bayes, Linear Discriminant Analysis, Quadratic Discriminant Analysis and Decision Trees. The database consisted of 2179 military personnel, divided between those who requested discharge and those who remained in service. The analysis identified significant sociodemographic and structural variables, such as age, specialty, commendations, and evaluation grade. The Logistic Regression (LR) model showed the best performance, allowing the calculation of a probability of departure index, with practical utility for human resources management. Despite limitations, such as low sensitivity and the use of a limited set of variables, the study offers important contributions to talent retention in the Air Force and suggests the inclusion of more variables and advanced approaches in future research.

Palavras-Chave

Retenção, Rotação, Métodos de *Machine Learning*

Keywords

Retention, Rotation, Machine Learning Methods

Índice

Introdução	9
1.Revisão da Literatura	13
1.1 Retenção de Talento.....	13
1.2 Rotação voluntária (<i>Turnover</i>).....	14
1.3 Saídas Voluntárias das Forças Armadas	16
2. Objetivos e Metodologias	20
2.1 Objetivos.....	20
2.2 Metodologia	21
2.3 Dados	27
2.4 Variáveis.....	28
3. Apresentação e discussão dos resultados.....	31
3.1 Análise exploratória dos dados	31
3.2 Seleção do modelo	35
3.2.2 Aperfeiçoamento do modelo	41
Conclusão	53
Referências bibliográficas	58
Apêndices	64
Apêndice 1 – Matriz de confusão 5 métodos	
Apêndice 2 – Gráficos das estatísticas descritivas.....	
Apêndice 3 – Resultado das variáveis do modelo inicial (21 variáveis).....	
Apêndice 4 – Resultado das variáveis do modelo RL 1.0	

Índice de Tabelas

Tabela 1 – Identificação Variáveis.....	29
Tabela 2 – Estatísticas das variáveis quantitativas	32
Tabela 3 – Contabilização de outliers por variáveis e status de Abates	34
Tabela 4 – Métricas de performance do modelo	37
Tabela 5 – Resultados matriz de confusão dos métodos	38
Tabela 6 – Interpretação da Pontuação AUC	40
Tabela 7 – Resultados RL antes e depois da seleção Stepwise	42
Tabela 8 – Resultado das variáveis do modelo após a seleção Stepwise.....	43
Tabela 9 – Resultados matriz de confusão RL original vs. RL 1.0.....	47

Índice de Gráficos

Gráfico 1 – Evolução Efetivos Quadros Permanentes	10
Gráfico 2 – Abates aos QP da FA 2014-2023	11

Índice de Figuras

Figura 1 – Método de Estudo de Caso	22
Figura 2 – Processo CRisp- DM	23
Figura 3 – Aprendizagem Supervisionada	25
Figura 4 – Distribuição das variáveis qualitativas	31
Figura 5 – Distribuição das variáveis quantitativas	33
Figura 6 – Matriz de confusão	35
Figura 7 – Representação de uma curva ROC.....	39
Figura 8 – ROC Curves dos métodos.....	40
Figura 9 – Matriz confusão RL 1.0.....	47
Figura 10 – Representação das variáveis significativas do modelo RL 1.0.....	48

Lista de abreviaturas

1SAR – Primeiro-Sargento

2SAR – Segundo-Sargento

ABST - Abastecimento

ADMAER- Administração Aeronáutica

AD – Árvores de Decisão

ALF - Alferes

CAP - Capitão

CMI - Construção e Manutenção de Infraestruturas

COR - Coronel

ENGAED - Engenheiros de Aeródromos

ENGAER - Engenheiros Aeronáuticos

ENGEL- Engenheiros Eletrotécnicos

FA – Força Aérea

FAV – Ficha de Avaliação

FFAA – Forças Armadas

GRH – Gestão Recursos Humanos

JUR – Juristas

LDA - *Linear Discriminant Analysis*

MAJ - Major

MARME - Mecânicos de Armamento e Equipamento

MED – Médicos

MELECA - Mecânicos de Eletrônica

MELECT - Mecânicos de eletricidade

MELIAV - Mecânicos de Eletricidade e Instrumentos de Avião

MMA - Mecânicos de Material Aéreo

MMT - Mecânicos de Material Terrestre

MUS - Banda e Fanfarra – Músicos

NAV – Navegadores

NB – *Naive Bayes*

OPCART - Operadores de Circulação Aérea e Radaristas de Tráfego

OPCOM - Operadores de comunicações

OPINF - Operadores de Informática

OPMET - Operadores de Meteorologia

OPRDET - Operadores Radaristas de Detecção
OPSAS - Operadores de Sistemas de Assistência e Socorros
PA - Polícia Aérea
PAPSM – Plano de Ação para a Profissionalização do Serviço Militar
PILAV – Pilotos Aviadores
PSI – Psicólogos
QP – Quadros Permanentes
QDA - *Quadratic Discriminant Analysis*
RC – Regime de Contrato
RH – Recursos Humanos
RL – Regressão Logística
SAJ – Sargento-Ajudante
SAS - Secretariado e Apoio dos Serviços
SCH – Sargento-Chefe
SMOR – Sargento-Mor
TABST - Técnicos de Abastecimento
TCOR – Tenente-Coronel
TEN - Tenente
TINF - Técnicos de Informática
TMAEQ - Técnicos de Manutenção de Armamento e Equipamento
TMI - Técnicos de Manutenção de Infraestruturas
TMMA - Técnicos de Manutenção de Material Aéreo
TMMEI - Técnicos de Manutenção de Material Eletrotécnico
TMMT - Técnicos de Manutenção de Material Terrestre
TOCART- Técnicos de Operações de Circulação Aérea e Radar de Tráfego
TOCC - Técnicos de Operações de Comunicações e Criptografia
TODCI - Técnicos de Operações Detecção e Conduta de Interceção
TOMET- Técnicos de Operações de Meteorologia
TPAA - Técnicos de Pessoal e Apoio Administrativo
TS - Técnicos de Saúde
TSE – Tempo Serviço Efetivo

Introdução

Atualmente, um dos grandes desafios que as organizações enfrentam é a retenção do seu capital humano. De acordo com Chiavenato (2014) e Souza et al (2022), tendo em consideração que as pessoas dão vida às organizações e que permitem que estas sobrevivam no meio social, é importante conseguir manter nos seus quadros este ativo organizacional.

Esta matéria tem vindo a ser considerada de extrema relevância para as organizações, já que a saída voluntária de um colaborador implica, necessariamente, custos acrescidos para a organização. Assim, quando o colaborador sai leva consigo todo o conhecimento e experiência adquiridos, sendo que a sua substituição irá acarretar custos ao nível do recrutamento, formação e qualificação (Meirinhos & Barreto, 2018) e (Nunes et al., 2021).

Para colmatar esta situação, nos últimos anos as organizações têm procurado compreender este fenómeno através de estudos, identificando fatores-chave que influenciam esta situação, por forma a conseguir compreender e mitigar as saídas voluntárias das organizações (Halim et al., 2020).

Neste contexto, as Forças Armadas (FFAA) não são exceção. A preocupação com a saída de militares dos quadros permanentes (QP) tem sido transversal aos três ramos das FFAA, Marinha, Exército e Força Aérea (FA), tendo sido reconhecido pelo Chefe do Estado-Maior-General das Forças Armadas, General Nunes da Fonseca, que "desde 2010 temos vindo a reduzir drasticamente os nossos efetivos. Estamos drasticamente abaixo dos efetivos autorizados e, portanto, só podemos dizer que há que lutar por ter mais efetivos" (Lusa, 2023b, para. 4).

Para esta redução drástica de efetivos tem contribuído um aumento significativo no número de abates¹ de militares dos QP das FFAA, sendo estes “cidadãos que, tendo ingressado voluntariamente nas Forças Armadas, prestam serviço profissional firmado em vínculo definitivo, designado por nomeação, constituindo fator da afirmação e perenidade dos valores da instituição militar” (Decreto-Lei n.º 90/2015 Estatuto dos Militares das Forças Armadas (EMFAR), 2015, p. 3218), que não tem conseguido ser compensado com a entrada de novos militares.

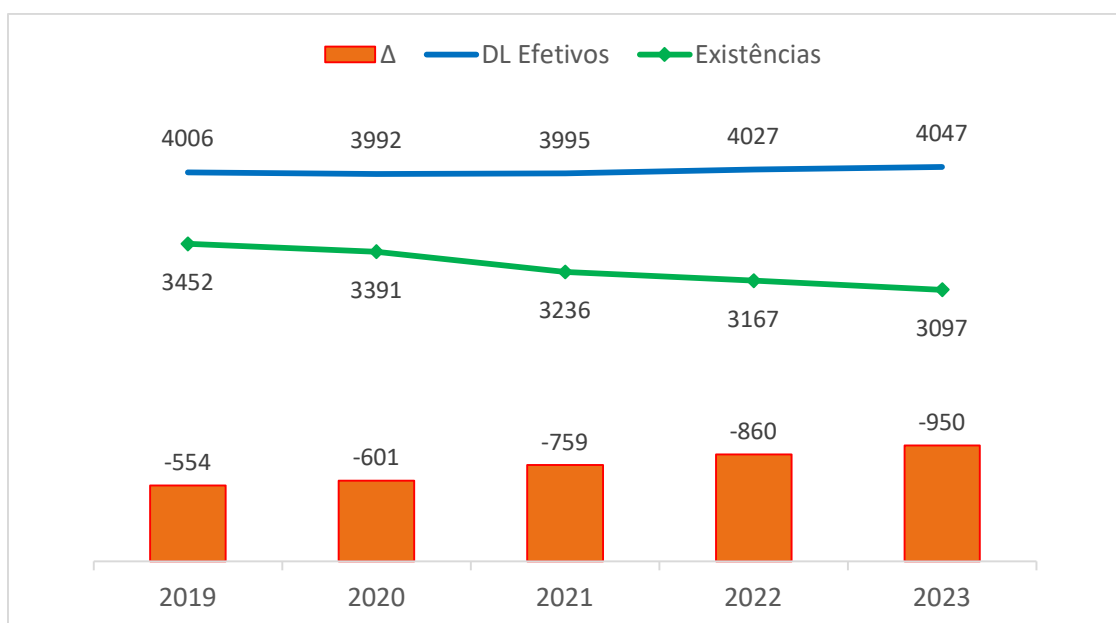
Desta forma, sendo as FFAA um elemento preponderante para a segurança do Estado e um “fator de projeção do prestígio internacional de Portugal” (Conceito Estratégico Defesa Nacional, 2013, p. 1982), é fundamental garantir a estabilidade dos seus quadros para o correto funcionamento do Estado.

¹ O abate aos QP está definido no artigo 171.º do EMFAR e implica uma desvinculação à organização castrense antes da carreira completa, significando uma saída não programada para a organização.

Seguramente, foi por esse motivo que, em 2019, foi criado o Plano de Ação para a Profissionalização do Serviço Militar (PAPSM), revisto em 2023, com o objetivo de “valorizar a profissão militar, garantindo um recrutamento eficiente e mantendo o efetivo existente nas Forças Armadas e promovendo a transição dos militares contratados para a vida civil”. Nesse sentido, o PAPSM prevê quatro eixos com as suas respectivas áreas de intervenção, sendo um deles o da retenção que, entre outras medidas, inclui a produção “de conhecimento científico sobre as saídas das Forças Armadas” (Plano De Ação Para a Profissionalização Do Serviço Militar (PASPM), 2023, p. 15).

No caso específico da FA, para o General João Cartaxo Alves, Chefe do Estado Maior deste Ramo, “o problema não reside em recrutar, já que este ramo se tem revelado atrativo para os jovens, mas “o que se passa é uma questão de competitividade com o mercado” (Lusa, 2023a, para. 2).

Gráfico 1 – Evolução Efetivos Quadros Permanentes



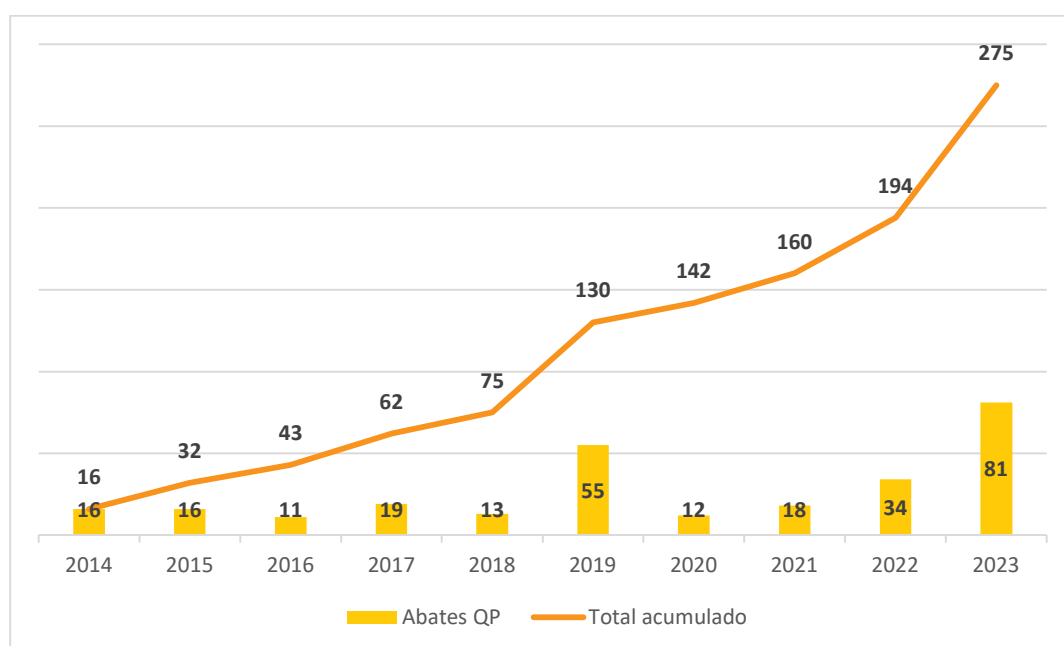
Fonte: Autoria própria, dados Força Aérea (2024)

No Gráfico 1 pode-se observar a evolução do efetivo dos militares dos QP na FA (existências), nos últimos 5 anos, comparados com o efetivo autorizado pelo Ministério da Defesa Nacional, através do denominado Decreto-Lei de Efetivos (DL Efetivos), e a respetiva diferença (delta).

Desta forma, fica evidenciado no Gráfico 1, que nos últimos cinco anos, a Força Aérea sofreu uma redução de cerca de 10% no seu efetivo de militares dos QP, indicando que o ramo enfrenta uma das situações de escassez de pessoal mais graves das últimas décadas (Alves, 2024).

Um dos fatores que tem contribuído para esta redução é o aumento das saídas não planeadas de militares dos QP (Alves, 2024). Embora não seja um problema recente, nos últimos anos tem-se registado um aumento significativo destas saídas, e, por não serem previstas, não têm sido compensadas pela entrada de novos militares, como podemos observar no Gráfico 2.

Gráfico 2 – Abates aos QP da FA 2014-2023



Fonte: Autoria própria, dados Força Aérea (2024)

A retenção de militares tem sido uma preocupação evidente para este ramo, sendo incluída como uma das linhas de ação na sua Diretiva Estratégica para o triénio 2022-2025. Neste sentido, é fundamental “rever o modelo de captação, formação e retenção de militares, de modo a adaptar a estrutura organizacional às exigências da condição militar, em alinhamento com os valores da instituição” (Força Aérea, 2022, p.32).

Com base nestas premissas, o tema deste trabalho de dissertação para a obtenção do grau de mestre em Ciência de Dados para Empresas, da Escola Superior de Ciências Empresariais do Instituto Politécnico de Setúbal é: **A Retenção de Militares dos Quadros Permanentes na Força Aérea – Os dados como fonte de informação e previsão.**

A pergunta de partida para esta investigação é: *será possível prever as saídas dos militares dos Quadros Permanentes da Força Aérea com base nos dados disponíveis sobre eles?*

Assim, o objetivo geral deste trabalho é estudar a problemática das saídas voluntárias (abates) dos militares dos QP da FA, utilizando os dados existentes nos sistemas de informação de recursos humanos da instituição. Como objetivos específicos, pretende-se identificar os indicadores (variáveis) presentes nesses sistemas que possam influenciar as saídas e definir um índice de risco de saída voluntária dos militares dos QP.

Serão utilizadas técnicas de inteligência artificial, mais precisamente de *machine learning*, particularmente os modelos de classificação.

Em termos metodológicos, esta investigação irá utilizar a metodologia de estudo de caso, uma vez que se pretende aprofundar a temática dentro da FA, e a metodologia de *CRoss Industry Standard Process for Data Mining* (CRISP-DM), mais utilizada para projetos de análise e ciência de dados.

O presente trabalho de investigação está estruturado em cinco capítulos: o primeiro capítulo, que consiste na introdução à temática; o segundo, dedicado à revisão da literatura; o terceiro, onde são apresentados os objetivos e a metodologia; o quarto, que aborda a apresentação e discussão dos resultados; e, por fim, o quinto capítulo, onde se apresentam as conclusões.

1.Revisão da Literatura

Neste capítulo, serão abordados três conceitos essenciais para uma melhor compreensão do objeto de estudo e um aprofundamento do tema em investigação: a retenção de talentos, a rotação de pessoal, ou *turnover*, e as saídas voluntárias nas FFAA.

1.1 Retenção de Talento

Vivemos numa era em que a retenção de colaboradores talentosos tem sido uma das maiores preocupações para as organizações (Halim et al., 2020). As organizações competem por ter os melhores recursos humanos, tornando-se assim responsável dos profissionais de Gestão de Recursos Humanos (GRH) encontrar as melhores estratégias e ferramentas para reter esses profissionais pelo maior tempo possível (Munish & Agarwal, 2017).

As organizações de sucesso perceberam que só conseguem continuar a crescer e a prosperar se conseguirem obter o máximo de retorno do investimento efetuado nos seus colaboradores (Chiavenato, 2014).

Mas porque é que a retenção de talento é tão importante para as organizações atualmente? Para Meirinhos & Barreto (2018), existem três fatores essenciais para se perceber esta questão: primeiro porque a retenção de talentos implica manter nos quadros da organização os profissionais mais produtivos e mais competentes; segundo porque quando um colaborador sai da organização leva consigo toda experiência e conhecimento adquiridos ao longo dos tempos; e terceiro porque existem custos associados à substituição do profissional que sai, nomeadamente no processo de recrutamento, formação e adaptação.

Esta ideia é também defendida por Lee & Mitchell (1994) e Nunes et al. (2021), que destacam o impacto significativo da perda de um profissional talentoso, que exigiu tempo e recursos para ser desenvolvido, e que quando sai leva consigo toda a experiência profissional adquirida.

É por isso que Mazlan & Jambulingam (2023) defendem que o foco da Organização na retenção de talentos deve estar nas atividades e iniciativas para atrair e reter os colaboradores mais talentosos.

As organizações devem intensificar o investimento em estratégias de retenção de talentos, uma vez que o papel desses profissionais é fundamental para o seu sucesso e para atingir os objetivos propostos (Nunes et al., 2021).

Por outro lado, quando se fala em Gestão de Pessoas, as organizações devem não só conseguir recrutar e formar talentos, como também conseguir mantê-los satisfeitos e comprometidos com a organização, por forma a evitar a sua saída prematura. Por isso, recrutar, fortalecer, utilizar e reter colaboradores com talento, é hoje um dos maiores desafios das organizações (Chiavenato, 2014).

Desta forma, segundo Nunes et al.(2021), as organizações que investem na retenção de talentos estão a contribuir para a diminuição da rotação dos colaboradores, ou *turnover*, que é outra das dimensões que mais as preocupa atualmente.

Um profissional altamente qualificado que não veja reconhecido o seu valor, irá certamente procurar esse reconhecimento noutra organização (Meirinhos & Barreto, 2018).

Em suma, a temática da retenção de talentos está cada vez mais presente na agenda das organizações, uma vez que os custos inerentes a uma saída precoce de um colaborador com talento, podem representar um enorme prejuízo, nem que seja pelo custo associado à sua substituição. Assim, as organizações têm procurado entender melhor esta dimensão, desenvolvendo estratégias que favoreçam a retenção do talento, e consequentemente, minimizem o *turnover*. Este conceito será aprofundado a seguir.

1.2 Rotação voluntária (*Turnover*)

A rotação de colaboradores é a outra face da retenção de talentos. Enquanto a retenção tem como preocupação a definição de estratégias para entender como reter os colaboradores mais talentosos, a rotação centra-se nas saídas voluntárias dos colaboradores e nos motivos pelos quais elas aconteceram. Em termos práticos, pode-se definir rotação como o tempo que os trabalhadores ficam numa determinada organização (Mafini & Dubihlela, 2013).

Na verdade, a decisão de um colaborador deixar a organização tem intrigado os estudiosos organizacionais e os gestores de RH há muito tempo (Mafini & Dubihlela, 2013). Para Lee & Mitchell (1994), a análise da rotação voluntária dos trabalhadores é, para as organizações, uma ferramenta essencial de gestão, enquanto que os investigadores é uma preocupação com o estudo das dimensões que levam um colaborador a sair voluntariamente da organização.

Dessa forma, de acordo com Souza et al. (2022), um *turnover* elevado pode provocar problemas sérios nas organizações, como a “perda de memória institucional”. Para estes autores, que estudaram a situação de *turnover* numa multinacional brasileira, a saída não deve ser atribuída apenas ao indivíduo, mas também a alterações organizacionais que afetam o grupo e que podem levar à saída coletiva dos colaboradores.

Para Mehl et al. (2022), a rotação de colaboradores é um dos principais problemas nas cadeias de abastecimento, sendo reconhecida pelos seus executivos como o maior desafio operacional, afetando diretamente as operações e comprometendo os investimentos cada vez mais elevados no recrutamento e formação dos colaboradores. Para além disso, surge uma nova situação que é o desgaste dos colaboradores que ficam na organização e que precisam de cumprir os objetivos com um menor número de pessoas.

Halim et al. (2020), ao estudarem os motivos pelos quais os colaboradores deixam voluntariamente uma organização, destacam que para além do baixo salário, e de uma oferta

de carreira mais atrativa por parte de outra organização, a cultura e o ambiente organizacional, como a intimidação no local de trabalho, são fatores importantes que influenciam a decisão do trabalhador em sair da organização.

Portanto, de acordo com Mehl et al. (2022), é essencial identificar quais as causas principais de saída, tendo em consideração que em cada etapa do ciclo de vida de um trabalhador o seu compromisso organizacional é diferente, e que os colaboradores que desempenham a mesma função podem apresentar motivos de saída diferentes. Neste contexto, as novas técnicas de inteligência artificial podem ajudar a compreender melhor este fenómeno.

Uma das teorias mais conhecidas sobre o *turnover* é a *Unfolding Model of Voluntary Turnover* desenvolvida por Thomas Lee e Terence Mitchell em 1994. Segundo estes autores, as pessoas são constantemente bombardeadas com informações que podem potenciar mudanças de comportamento. Este modelo de “desdobramento” procura retratar a rotatividade dos trabalhadores como um processo complexo, no qual os colaboradores ponderam os seus sentimentos, a sua situação pessoal e o seu ambiente de trabalho tomando, ao longo do tempo, a decisão de permanecer ou sair da organização.

Este modelo é caracterizado por cinco componentes: o primeiro diz respeito aos **Choques**, que podem ser definidos como os eventos que influenciam a decisão do indivíduo em permanecer, ou não, na organização; o segundo componente é designado por **Planos** e caracteriza-se pelas respostas que os colaboradores já desenvolveram para enfrentar certos choques; o terceiro é denominado por **Procura e Avaliação de Alternativas**, sendo o processo em que, após um choque, o colaborador começa a considerar as alternativas ao seu trabalho atual; o quarto componente é designado por **Satisfação no Trabalho e Compromisso Organizacional** e traduz os sentimentos de satisfação com o trabalho e o compromisso organizacional do colaborador; e por fim, o quinto componente que se designa por **Caminhos de Decisão** e que define as diferentes trajetórias que o colaborador pode seguir quando decide sair.

Existem quatro caminhos possíveis para um indivíduo que decide sair da organização, o primeiro traduz-se numa decisão imediata após um choque significativo, o segundo numa decisão baseada numa avaliação cuidada das alternativas existentes, o terceiro é baseado num choque que leva à insatisfação e que, por sua vez, a uma procura ativa de novas oportunidades de trabalho, e o quarto caminho é a decisão de sair baseada numa insatisfação que foi sendo criada ao longo do tempo em que o colaborador esteve na organização (Hom et al., 2017).

Este modelo tem sido validado e aceite ao longo dos anos, em que um dos caminhos mais simples deste modelo é aquele em que a insatisfação com o atual trabalho leva os colaboradores a considerarem a mudança de emprego, sendo que quando as oportunidades externas são apelativas pode gerar a intenção de mudança (Sims et al., 2024).

Para Hom et al. (2017), este modelo é de extrema importância devido à sua capacidade de capturar a complexidade do processo de decisão da saída voluntária, uma vez que possui uma estrutura robusta de análise tanto dos fatores imediatos (choques) como dos fatores contínuos (e.g. satisfação no trabalho) que podem influenciar a rotatividade.

Este modelo é também importante para compreender este fenómeno, uma vez que poderá ser utilizado tanto para fins de pesquisa, na análise de padrões de rotatividade e previsão de futuras saídas, como na prática organizacional no desenvolvimento de estratégias de retenção, após identificação dos motivos da saída (Hom et al., 2017).

Segundo Mehl et al. (2022), é possível transformar o *turnover* em retenção através de técnicas de análise avançada, como a regressão ou a classificação. O que se pretende é tratar a doença, que neste caso é a falta de compromisso organizacional, e não o sintoma que é a saída.

Resumidamente, pode-se afirmar que, atualmente, a rotação voluntária de trabalhadores é uma grande preocupação das organizações, e que estas têm procurado entender melhor este fenómeno através de estudos e análises. No entanto, verifica-se que a análise e a interpretação deste problema são complexas e que depende muito da experiência vivida pelo colaborador enquanto indivíduo, mas também pode ser influenciada pelo contexto organizacional e pelo que a organização oferece.

1.3 Saídas Voluntárias das Forças Armadas

O problema da saída voluntária de colaboradores e a consequente falta de retenção dos mesmos, não tem afetado somente a sociedade civil. A nível mundial, as FFAA também têm sentido este problema sendo que, segundo Sims et al. (2024), a retenção de militares tem sido considerada, nos últimos tempos, como de elevado interesse. Este sentimento é partilhado por Gultekin et al. (2012), que defendem que altos níveis de atrição nas FFAA significam um elevado desperdício de investimento e uma redução na sua eficácia.

Mafini & Dubihlela (2013) estudaram as saídas voluntárias dos controladores aéreos da Força Aérea Sul-Africana e concluíram que os fatores que incentivam as pessoas a permanecer na organização incluem condições de trabalho justas, a realização de tarefas desafiantes e a promoção de relações harmoniosas entre os colaboradores.

Por outro lado, McMahon & Bernard (2018) analisaram as saídas voluntárias das FFAA Norte-Americanas e concluíram que muitas vezes a decisão de mudar de emprego é motivada pelo facto da organização não conseguir acompanhar o desenvolvimento dos seus colaboradores. Concluíram ainda que os incentivos financeiros, por si, não são ferramentas eficazes para recrutar e reter os militares, sendo necessário criar medidas e incentivos mais radicais para esse efeito, como a revisão do passado criminal dos jovens para avaliar a sua aptidão para uma carreira militar, ou a melhoria do estilo de vida familiar e profissional dos militares.

Também Bosdotter et al. (2022) analisaram o fenómeno das saídas voluntárias de pilotos da Força Aérea Sueca, que não tem sido compensada pelo número de entradas. Os autores realizaram uma série de entrevistas a pilotos militares, que saíram antes do fim da carreira, e concluíram que as razões apresentadas são significativamente diferentes e complexas, o que poderá indicar que a melhor forma de os manter pode variar de indivíduo para indivíduo. Assim, sugerem que pode existir mais do que uma perspetiva teórica para explicar os motivos das saídas dos pilotos.

De igual modo, Gultekin et al, (2012) também estudaram a saída voluntária de pilotos da Força Aérea Romena, através de modelos de regressão, e concluíram que não existe um caminho claro para prever a decisão de uma pessoa de sair da organização. No entanto, existem benefícios que a organização poderá retirar deste estudo, nomeadamente na gestão dos recursos humanos, ao constatar que a aviação civil, o nível educacional e o percurso de carreira podem influenciar a decisão de saída dos pilotos.

Numa perspetiva mais psicológica, Congard et al. (2022) estudaram a preponderância dos traços de personalidade na decisão de rotatividade voluntária, tendo concluído que elevados níveis de neuroticismo e de stress aumentam o risco de rotatividade, e que elevados níveis de agradabilidade, quando associados a más relações no trabalho, favorecem uma maior rotação. Para os autores, a maioria dos estudos tem procurado explicações individuais ou contextuais para este problema. No entanto, o estudo que realizaram analisou a interação entre os traços de personalidade e a experiência subjetiva de trabalho para procurar compreender a rotação voluntária.

Em Portugal, a problemática das saídas voluntárias nas FFAA também tem vindo a ser estudada nos últimos anos.

Martins (2014), ao estudar a problemática do *turnover* dos militares do Regime de Contrato (RC), concluiu que as causas da saída voluntária estão associadas a aspetos ligados com o projeto profissional, bem como à quebra das expectativas iniciais. Assim, estes militares ao cumprir o serviço militar substituíram o vetor emocional pelo racional, percecionando a organização militar uma opção pouco sustentável.

Também Rijo et al. (2018) estudaram a mesma temática e concluíram que a saída precoce dos militares em RC deve-se essencialmente a características instrumentais da organização, como o baixo vencimento ou a perceção de alternativas de trabalho mais aliciantes, bem como aos indicadores macroeconómicos de emprego que se traduzem em oportunidades de trabalho mais vantajosas na sociedade civil. Desta forma, quando os militares conjugam estes fatores, aumenta a sua perceção desfavorável sobre a qualidade de vida nas FFAA, contribuindo para a insatisfação no trabalho e para a diminuição do compromisso organizacional. Para os autores, as saídas voluntárias de militares RC resultam na perda de

capital humano especializado numa determinada área de atuação, bem como de conhecimento e especialização adquiridos ao longo dos anos.

Do mesmo modo, Santos (2021) também estudou esta problemática e concluiu que o desenvolvimento profissional, associado à gestão de qualificações ou competências, as condições de trabalho, de habitabilidade e de apoio social, bem como as condições financeiras têm um impacto direto na retenção dos militares RC.

Já na perspetiva da saída voluntária dos militares dos QP, Branco (2022) refere que a insatisfação dos militares do Exército está principalmente relacionada com a questão salarial. Das entrevistas realizadas, ficou claro que as alternativas laborais na sociedade civil ofereciam melhor remuneração, especialmente quando comparadas com o n.º de horas de trabalho. Concluiu ainda que a GRH é uma componente imprescindível na gestão da satisfação dos militares podendo ter impacto direto na intenção de saída.

No caso da Marinha, Neves (2023) estudou as saídas extemporâneas dos militares dos QP e do RC, através da análise dos questionários de saída aplicados a esses militares, e concluiu que os militares dos QP identificaram três principais dimensões para deixarem a organização: “Remuneração e incentivos”, a “Carga de trabalho” e o “Equilíbrio entre o trabalho e vida pessoal”. De igual modo, concluiu ainda que a “decisão de sair da instituição não resulta apenas de um motivo, mas sim de uma multiplicidade de razões que se conjugam no momento da decisão” Neves (2023).

Por outro lado, Machado & Fachada (2018) estudaram a eficácia da retenção dos militares dos QP da FA, em resposta à política de indemnizações em vigor, e concluíram que o compromisso organizacional é o principal motivo pelo qual os militares desejam permanecer na Instituição. Além disso, apresentaram medidas concretas para a retenção desses militares, que incluem a revisão do sistema de recompensas, a gestão moderada das colocações e destacamentos e a disponibilização de meios de suporte, justiça e comunicação organizacional.

Em suma, pode-se afirmar que existe, de facto, uma preocupação crescente com a falta de retenção e consequentes saídas voluntárias dos militares das FFAA. Verifica-se que esta temática tem sido estudada nos últimos anos, tanto a nível internacional, como a nível nacional, com o intuito de compreender melhor este fenómeno.

No entanto, dos estudos apresentados, especialmente a nível nacional, observa-se que estes apresentam algumas limitações para o presente trabalho, uma vez que são circunscritos a militares do RC, ou, no caso dos estudos que analisaram os militares dos QP focarem-se apenas na perceção dos militares após a sua saída, sem procurar obter respostas junto dos militares que ainda se encontravam na organização.

2. Objetivos e Metodologias

Neste capítulo, serão discutidos os objetivos e as metodologias selecionadas para responder à pergunta de investigação. Também serão apresentados os dados e as variáveis selecionadas para este estudo.

2.1 Objetivos

De acordo com Chiavenato (2014), estamos na Era da Informação em que a gestão de pessoas deixou de ser apenas um desafio e passou a ser também uma vantagem competitiva entre as organizações.

Assim, conforme referido anteriormente, o principal objetivo desta investigação é estudar a problemática das saídas voluntárias dos militares dos QP da FA.

As preocupações com as saídas voluntárias dos militares dos QP não são exclusivas aos três ramos das FFAA, mas têm também sido uma questão de interesse para a tutela, já que “as pessoas são o pilar central da Defesa Nacional” (PAPSM, 2023, p1).

. Para a ex-Ministra da Defesa, Professora Dra. Helena Carreira, o “objetivo de consolidar o atual modelo de prestação de serviço militar centrado nas pessoas é fundamental. O trabalho aqui apresentado reforça a atenção dispensada às nossas principais prioridades: queremos atrair mais e melhor, para também podermos reter mais e melhor” (PAPSM, 2023, p1).

Recentemente, o atual Ministro da Defesa, Dr. Nuno Melo, também reconheceu este problema nas FFAA, afirmando que “temos hoje problemas de recrutamento e de retenção porque os salários são baixos, os suplementos não são adequados, o alojamento muitas vezes não existe ou é caro, muitas instalações não estão recuperadas, o equipamento poderia ser muito melhor e, a par disto, o que tem a ver também com os ex-combatentes, num estatuto que pode ser melhorado com outras garantias que o Estado deve assegurar” (Lusa, 2024, para. 11).

Desta forma, o Governo e os Chefes do Estado Maior dos vários Ramos das FFAA têm sinalizado, nestes últimos anos, a retenção como um dos problemas que afeta as FFAA.

No caso da FA, a situação tem vindo a agravar-se desde 2021 e tem provocado um desequilíbrio entre as entradas nos QP e as saídas previstas. Ou seja, dado que um pedido de abate é uma situação imprevisível, implica sempre um desequilíbrio nos efetivos (Alves, 2024).

Conforme mencionado anteriormente, o Gráfico 2 ilustra claramente esta situação. Nos últimos cinco anos, observou-se um acentuado aumento no número de abates dos militares dos QP, apenas atenuado pela COVID-19 nos anos de 2020 e 2021.

No entanto, tal como referem Munish & Agarwal (2017), a gestão de topo precisa de conhecer e compreender a perceção e a rotatividade dos seus colaboradores, uma vez que as suas

preferências e atitudes podem mudar rapidamente. Torna-se, assim, essencial reduzir a taxa de rotatividade adotando estratégias de retenção inovadoras e eficazes.

Por esta razão, é fundamental estudar a problemática da rotatividade dos militares dos QP da FA, a fim de estabelecer a estratégia ideal para mitigar esta situação, com base nos motivos que levam os militares a abandonar a instituição. No âmbito desta investigação, procurar-se-á estudar esta questão exclusivamente com base nos dados existentes nos sistemas de informação da área de Recursos Humanos da FA.

É importante salientar que o foco deste trabalho estará sempre na génese deste mestrado, o da ciência de dados, e nunca na perspetiva sociológica ou psicológica. Assim, serão analisados apenas os dados existentes nesses sistemas.

É com esta perspetiva que se define um dos objetivos específicos deste trabalho, o de tentar identificar os indicadores que poderão influenciar os abates aos QP da FA, utilizando os dados existentes nos sistemas de informação da área de Recursos Humanos da FA, gerados e armazenados ao longo da carreira dos militares.

Como refere Chiavenato (2014), os dados após serem trabalhados, transformam-se em informação. Assim, o objetivo é perceber se esses dados podem “contar uma história”, ou seja, se a evolução da carreira de um militar pode indicar se ele estará predisposto para pedir um abate aos QP.

A identificação de indicadores-chave que influenciam a retenção dos trabalhadores é uma tarefa crucial da investigação, dado que existem múltiplas razões pelas quais um colaborador pode optar por sair da organização (Halim et al., 2020).

Ao identificar as variáveis que poderão influenciar a intenção de pedido de abate, é possível estabelecer um segundo objetivo específico que é a definição de um índice de risco de saída associado a cada um dos militares dos QP da FA. Desta forma, pretende-se atribuir a cada um destes militares um índice de probabilidade de saída que permita antecipar um possível abate aos QP.

Este índice de risco poderá ser uma ferramenta útil na GRH, permitindo à organização implementar estratégias proativas de retenção e, assim, mitigar a rotatividade dos seus militares.

Em resumo, pretende-se responder à questão de partida já enunciada de prever o pedido de abate aos QP dos militares da FA com base nos dados existentes sobre eles.

2.2 Metodologia

Para o desenvolvimento deste trabalho optou-se por utilizar duas metodologias: uma mais clássica denominada de estudo de caso; e a outra centrada na natureza deste mestrado de ciência de dados, o CRISP-DM.

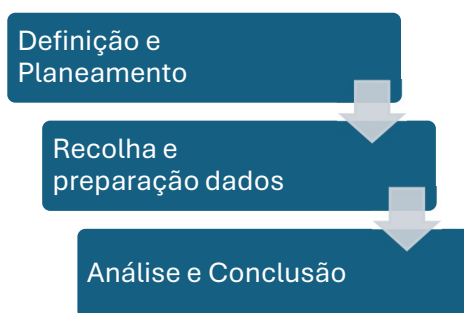
A escolha pela metodologia de estudo de caso justifica-se pelo facto de, segundo Yin, (2001), esta estratégia de investigação ser adequada quando se pretende responder a questões do tipo “como” e “porque”, com o foco em fenómenos da atualidade inseridos num contexto específico da vida real.

O estudo de caso pode ser considerado um meio eficaz para investigar o objeto de estudo no seu contexto real, utilizando diferentes fontes de dados, tanto qualitativas, quanto quantitativas, com o intuito de criação de conhecimento (Meirinhos & Osório, 2010).

Para Santos & Lima (2019), o estudo de caso deve ser entendido como um procedimento que visa estudar uma única entidade de estudo, seja um “indivíduo, a comunidade ou mesmo a nação”, através da recolha detalhada de informação sobre essa unidade.

Nesse sentido, o estudo de caso é um contributo singular para a compreensão de fenómenos organizacionais complexos (Yin, 2001) e uma estratégia poderosa para explicar contextos multifacetados, envolvendo variáveis igualmente complexas (Meirinhos & Osório, 2010).

Figura 1 – Método de Estudo de Caso



Fonte: adaptado de Yin (2001)

Como se pode observar na figura 1, a aplicação da metodologia do estudo de caso pode ser simplificada em três fases principais. Na primeira fase é definido o problema e é elaborado o plano de investigação que irá orientar o estudo. A segunda fase é dedicada à recolha e preparação dos dados, garantindo que os mesmos são adequados para a análise. Na terceira fase procede-se à análise dos dados e à formulação das respetivas conclusões.

Dado o âmbito restrito desta investigação, que se foca na análise dos abates aos QP dos militares da FA, considera-se que esta metodologia é a mais indicada para explorar o fenómeno em detalhe e fornecer *insights* baseados em dados. O estudo de caso permite aprofundar a compreensão deste fenómeno complexo e específico, ajustando-se ao contexto particular da FA.

Porém, tendo em consideração a natureza do mestrado em que se insere este trabalho de dissertação, e a natureza dos objetivos definidos, considerou-se também um método ligado à ciência de dados.

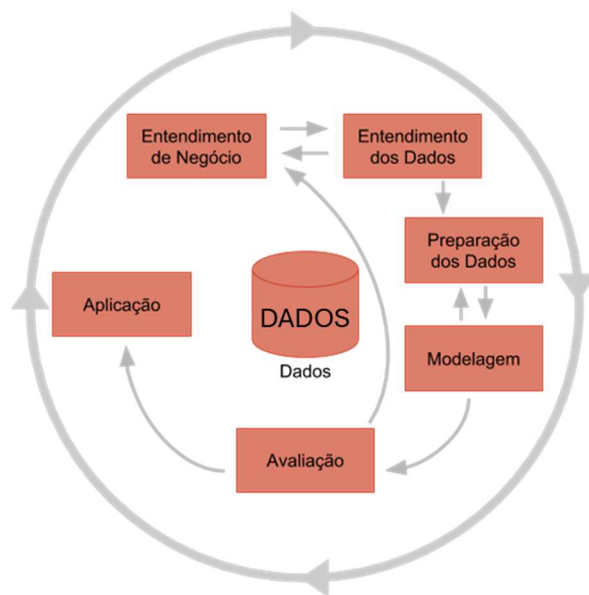
Segundo Saltz (2021), apesar de existirem muitos conceitos acadêmicos e organizacionais associados à ciência de dados, todos partilham o mesmo objetivo: gerar valor por meio da análise de dados. Também Martinez et al. (2021) defendem essa ideia, enfatizando de que a ciência de dados visa resolver problemas e obter conhecimento através da análise de dados. Entre as várias metodologias disponíveis para projetos de ciência de dados, optou-se pelo CRISP-DM, desenvolvido na década de 90 do século passado e utilizado desde então para definir o modelo de ciclo de vida de projetos que procuram o conhecimento através da utilização de base de dados (Saltz, 2021).

Segundo Martinez-Plumed et al. (2021), o CRISP-DM é um processo que se inicia com um objetivo claro e definido, e onde os dados recolhidos estão disponíveis para serem processados informaticamente.

Apesar dos autores considerarem que a ciência de dados ainda é um assunto recente, com diversas questões em aberto sobre a sua natureza e metodologia, a realidade é que a CRISP-DM continua a ter um papel importante na definição de uma estratégia para a realização de projetos de análise de dados.

Segundo Hayat Suhendar & Widayani (2023), Martinez-Plumed et al. (2021), Martinez et al.(2021) e Saltz (2021), o processo de CRISP-DM pode ser visto como um ciclo constituído por 6 fases interativas, que ajudam no desenvolvimento e progresso dos projetos de análise de dados.

Figura 2 – Processo CRisp- DM



Fonte: Adaptado de Martinez-Plumed et al. (2021)

A figura 2 apresenta a dinâmica das 6 fases do CRISP-DM: na primeira fase é necessário existir o entendimento do negócio para perceber o problema a solucionar; a segunda fase corresponde ao entendimento dos dados com o intuito de descobrir que tipo de dados existem e sobre os quais a solução vai ser encontrada; na terceira fase os dados são preparados e trabalhados para se obterem melhores resultados; na quarta fase é efetuada a modelação dos dados, ou seja, é aplicado um algoritmo de *machine learning* aos dados, de forma a tentar obter algum padrão; na quinta fase é efetuada uma avaliação dos resultados obtidos na fase anterior, de maneira a se obter algum grau de confiança nos mesmos; por fim, na sexta fase é efetuada a aplicação do modelo de *machine learning*, na tentativa de prever resultados futuros.

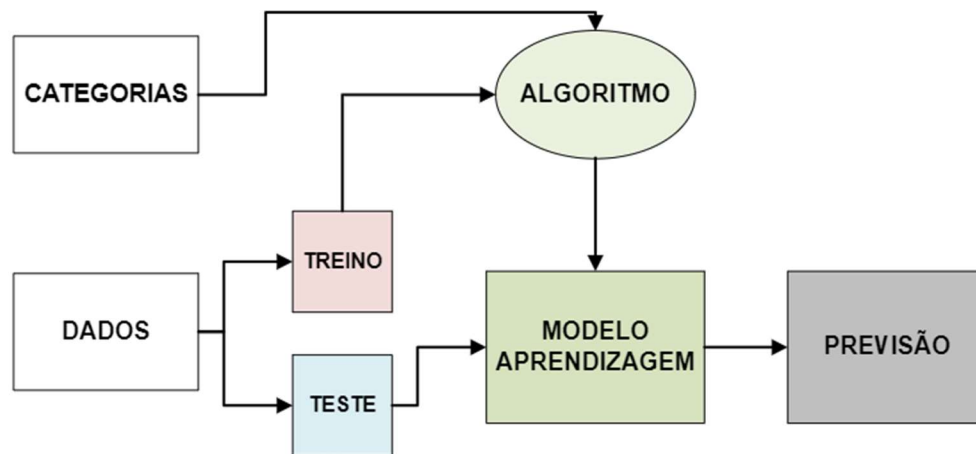
Tal como foi referido, será na quarta fase do processo onde serão aplicadas técnicas de inteligência artificial, mais precisamente de *machine learning*, na tentativa de criar um modelo que explique os dados (Hayat Suhendar & Widayani, 2023).

Assim, de acordo com Mohamed (2017), o conceito de *machine learning* não deve apenas ser visto como um armazenamento de grandes quantidades de dados, mas como parte integrante da inteligência artificial, que utiliza os computadores para realizar tarefas destinadas aos humanos.

Por outro lado, de acordo com Nasteski (2017), existem diversos tipos de algoritmos de *machine learning* que estão organizados de acordo com o resultado desejado. No âmbito desta investigação, irá ser utilizada a aprendizagem supervisionada, recorrendo a métodos de classificação que procuram mapear as variáveis de entrada com a da saída desejada, ou seja, o principal objetivo deste tipo de aprendizagem é aprender o caminho entre a entrada e

a saída, quando o valor deste último é conhecido e fornecido por um supervisor (Mohamed, 2017).

Figura 3 – Aprendizagem Supervisionada



Fonte: Adaptado de Nasteski (2001) e Singh et al. (2016)

Na figura 3 pode-se observar a aplicação da aprendizagem supervisionada num projeto de ciência de dados, onde os dados iniciais são divididos em duas amostras, uma que permitirá treinar o algoritmo e a outra que irá permitir avaliar e validar o modelo de aprendizagem. As categorias representam o valor final observado e que será fornecido ao algoritmo. Na prática, os dados de treino procuram encontrar as variáveis que explicam o resultado (categorias), enquanto os dados de teste irão permitir fazer a simulação do resultado e compará-lo com a realidade. No final, o objetivo é encontrar um modelo de aprendizagem que permita efetuar previsões baseadas numa nova base de dados.

No caso concreto desta investigação, as categorias serão compostas por dois valores, abate ou não abate, e o modelo de aprendizagem procurará encontrar uma forma de prever se um determinado militar irá pedir, ou não, abate aos QP. Considerando que se está perante uma variável qualitativa nominal (binária), utilizar-se-á os modelos de classificação e verificar qual o que produz melhores resultados para o nosso problema.

Assim, tal como preconizado por (Marvin et al., 2021; Nasteski, 2017; Singh et al., 2016; Yedida et al., 2024), o conjunto de dados será submetido a cinco modelos distintos de algoritmos de classificação: Regressão Logística (RL), *Naive Bayes* (NB), Análise Discriminante Linear (LDA do inglês *Linear Discriminant Analysis*), Análise Discriminante Quadrática (QDA, do inglês *Quadratic Discriminant Analysis*) e Árvores de Decisão (AD), que de acordo com James et al. (2023) podem ser definidos nestes termos:

- **RL:** A regressão logística é um método estatístico de classificação utilizado para modelar a relação entre uma variável dependente binária, assumindo dois valores

possíveis, e uma ou mais variáveis independentes (ou preditoras). Particularmente, utiliza a função logística para modelar a probabilidade de a variável assumir um dos dois valores possíveis, dado um conjunto de p características. A probabilidade estimada é convertida numa das categorias (0 ou 1) aplicando-se u, limiar, geralmente de 0,5.

- A LDA, a QDA e o classificador NB são métodos estatísticos de classificação, baseados no Teorema de Bayes, que visam estimar as probabilidades de uma observação pertencer a cada uma de k classes, dado um conjunto de p características, classificando a observação na classe com maior probabilidade. Contudo estes três métodos aplicam o Teorema de Bayes com diferentes abordagens e considerando diferentes suposições:

- **NB:** O termo "naive" (ingênuo) refere-se à suposição simplificadora de que todas as características, as variáveis preditoras, são condicionalmente independentes entre si, dado a classe C . A simplicidade do classificador *Naive Bayes* torna-o rápido e eficaz, mas a suposição de independência entre as variáveis preditoras nem sempre é realista, o que pode comprometer o desempenho em contextos onde as variáveis preditoras estão correlacionadas;
- **LDA:** Neste método assume-se que as variáveis preditoras seguem uma distribuição normal multivariada em cada uma das classes, com uma matriz de covariância comum. A simplificação de se admitir que a matriz de covariância das variáveis preditoras são iguais para todas as classes, permite obter funções discriminantes lineares, que são combinações lineares das variáveis preditoras. É um método eficaz em problemas com duas ou mais classes, no entanto pode ser limitado quando as suposições de normalidade e de homogeneidade das variâncias não são verificadas;
- **QDA:** É um método semelhante à LDA, mas permite que matriz de covariância das variáveis preditoras seja diferente de classe para classe, resultando numa maior flexibilidade. Nesse caso, as funções discriminantes são quadráticas e permitem uma separação mais complexa entre as classes. Contudo, requer a estimativa de um maior número de parâmetros, o que pode ser uma desvantagem em conjuntos de dados de dimensão pequena;

- **AD:** É um método não paramétrico utilizado tanto para a classificação como para a regressão. Pretende-se dividir o espaço p dimensional das variáveis preditoras num conjunto de regiões distintas e que não se intersectam, de forma a otimizar um determinado critério. O conjunto das regras de divisão determinadas para segmentar o espaço das variáveis explicativas pode ser representado numa árvore, que é comporta por:

- Nó Raiz: O ponto de partida da árvore, onde a primeira decisão é feita com base na variável preditora que melhor divide os dados.

- Nós internos: Representam uma condição sobre uma variável preditora.
- Folhas (ou Nós Finais): Representam as classes finais.
- Ramos: Conectam os nós e representam as decisões baseadas nas respostas a cada condição.

Para cada nó, são avaliadas todas as variáveis preditoras e suas possíveis divisões. As métricas mais comuns para avaliar a qualidade de uma divisão são:

- Índice Gini: Usado para medir a pureza de um nó. Quanto menor o índice, mais puro o nó, indicando que o nó contém predominantemente observações de uma só classe.
- Entropia: Uma medida da incerteza, que é zero quando todas as observações pertencem à mesma classe.
- Ganho de Informação: A diferença entre a entropia do nó pai e a soma ponderada das entropias dos nós filhos após a divisão.

O nó é dividido na variável que maximiza o ganho de informação (ou minimiza o índice Gini e a entropia). O processo é repetido recursivamente para cada novo nó, criando uma estrutura de árvore. O crescimento da árvore para quando um critério de paragem é atingido: profundidade máxima da árvore, número mínimo de observações em um nó ou homogeneidade dos nós, onde todos as observações pertencem à mesma classe.

Para avaliar a eficácia dos métodos da classificação referidos anteriormente irá ser utilizada a matriz de confusão, uma vez que possibilita a obtenção de um conjunto de métricas que permite avaliar o desempenho de cada um deles (Chawla et al., 2020; James et al., 2023; Lantz, 2013; Mohamed, 2017). O método que apresentar melhores resultados nessas métricas será utilizado para prever os abates aos QP e criar o índice de saída.

A base de dados a ser utilizada será explicada a seguir, tanto no que diz respeito à amostra quanto às variáveis.

2.3 Dados

Este trabalho irá centrar-se exclusivamente nos pedidos de abate voluntários por parte dos militares, ao abrigo das alienas c) e d) do artigo 171º do EMFAR, que estabelecem que os militares podem requer o abate aos QP após um tempo mínimo de serviço efetivo na categoria de ingresso nos QP, ou mediante o pagamento de uma indemnização (EMFAR, 2015).

Para definir a composição da amostra a recolher, ou seja, quais serão os militares a incluir neste estudo, optou-se, em primeiro lugar, pela seleção dos militares que pediram o abate aos QP da Força Aérea e depois pela escolha dos militares que não o fizeram

Assim, para a seleção dos militares que pediram abate aos QP utilizou-se o Gráfico 2 e verificou-se que existe de facto uma alteração no paradigma das saídas voluntárias dos

militares do QP a partir de 2019. Verifica-se que entre 2014 e 2018 o número de abates nunca foi superior a 19, e a partir desse ano os números aumentaram exponencialmente, sendo apenas atenuados, previsivelmente, pelos efeitos do COVID-19 na sociedade portuguesa.

Ao analisar os últimos 10 anos de abates aos QP na Força Aérea verifica-se que entre 2014 e 2018 existiram 75 saídas e que nos cinco anos seguintes existiram 200 saídas, o que representa um aumento de quase 170%. Assim, serão apenas considerados 200 militares que pediram abate aos QP entre os anos de 2019 e 2023.

Foi também feita uma análise à data de incorporação destes 200 militares e verificou-se que 5 deles tinham incorporado há mais de 30 anos, pelo que se optou por excluí-los desta investigação por se encontrarem numa fase diferente da sua carreira.

Para a seleção dos militares que não pediram abate, foram selecionados aqueles que ingressaram nos QP na mesma altura dos que pediram, considerando que entraram todos com as mesmas condições e expectativas de carreira.

Desta forma, a base de dados final é composta por 2005 militares que não saíram da FA e 195 militares que solicitaram abate aos QP, perfazendo um total de 2200 militares.

Importa referir que estes 2005 militares dos QP representam cerca de 65% dos militares dos QP na efetividade de serviço a 31 de dezembro de 2023 (Gráfico 2).

Tal como referido na figura 3, a amostra será dividida em duas: a de treino e a de teste. Na primeira, os dados de treino serão utilizados para “ensinar” o algoritmo a construir o modelo de aprendizagem, enquanto que na fase de teste é aplicado o modelo de aprendizagem aos dados de teste por forma a classificá-los de acordo com os rótulos estabelecidos (Nasteski, 2017).

Nesta investigação irá ser utilizado o rácio comumente aceite de 70% para a amostra de treino e 30% para a amostra de teste (Mitchell, 1997; Vrigazova, 2021).

Por último, foram retirados os valores nulos da amostra, ficando assim a base de dados final constituída por 2176 militares, 1993 que não pediram abate e 183 que o fizeram, tendo como data de referência 31 de dezembro de 2023.

2.4 Variáveis

Depois de selecionada a amostra, procedeu-se à recolha e compilação das variáveis que compõem a base de dados que foi utilizada para a análise da problemática da saída voluntária dos militares dos QP da FA.

Desta forma, foram selecionadas 22 variáveis que se dividiram em dois grupos: sociodemográficas e estruturais. As primeiras identificam características do indivíduo e sobre as quais a organização não tem nenhuma influência, como o número de filhos ou o estado civil; as segundas dizem respeito a características organizacionais que vão sendo adquiridas

ao longo do tempo em que o militar presta serviço, como o número de condecorações ou o número de formações frequentadas.

Tabela 1 – Identificação Variáveis

Variáveis		Descrição	Tipo
Sociodemográficas	Sexo	Sexo do militar	Nominal
	Estado Civil	Estado Civil do militar com as seguintes categorias: - Casado; - Unido de Facto; - Divorciado; - Viúvo; - Solteiro;	Nominal
	N Filhos	Número de filhos do militar	Quantitativa
	Residência	Tipologia de área urbana (TIPAU) da freguesia de Residência do militar. “A TIPAU constitui uma nomenclatura territorial atualizada do grau de urbanização de Portugal mediante a classificação tripartida das freguesias do território nacional em “Áreas predominantemente urbanas (APU)”, “Áreas mediantemente urbanas (AMU)” e “Áreas predominantemente rurais (APR)”.” (Agência para a Modernização Administrativa, 2020).	Nominal
	Idade	Idade do militar	Quantitativa
	Habilitação	Habilitação literária do militar até 31DEZ2023 ou à data de abate aos QP, dividida em dois grupos: Ensino Superior e Ensino Secundário.	Ordinal
Estruturais	Abate	Indica se o militar pediu abate aos QP com as seguintes categorias: - Não - Sim	Nominal
	Categoria	Categoria do militar com as seguintes categorias: - Oficial; - Sargento.	Ordinal
	Especialidade	Quadro Especial do militar com as seguintes categorias: ABST - Abastecimento ADMAER- Administração Aeronáutica CHBM – Chefe da Banda Música CMI - Construção e Manutenção de Infraestruturas ENGAED - Engenheiros de Aeródromos ENGAER - Engenheiros Aeronáuticos ENGEL- Engenheiros Eletrotécnicos JUR - Juristas MARME - Mecânicos de Armamento e Equipamento MED – Médicos MELECA - Mecânicos de Eletrónica MELECT - Mecânicos de eletricidade MELIAV - Mecânicos de Eletricidade e Instrumentos de Avião MMA - Mecânicos de Material Aéreo MMT - Mecânicos de Material Terrestre MUS - Banda e Fanfarra – Músicos NAV - Navegadores OPCART - Operadores de Circulação Aérea e Radaristas de Tráfego OPCOM - Operadores de comunicações OPINF - Operadores de Informática OPMET - Operadores de Meteorologia OPRDET - Operadores Radaristas de Deteção OPSAS - Operadores de Sistemas de Assistência e Socorros PA - Polícia Aérea PILAV – Pilotos Aviadores PSI – Psicólogos SAS - Secretariado e Apoio dos Serviços TABST - Técnicos de Abastecimento TINF - Técnicos de Informática TMAEQ - Técnicos de Manutenção de Armamento e Equipamento TMI - Técnicos de Manutenção de Infraestruturas TMMA - Técnicos de Manutenção de Material Aéreo TMEL - Técnicos de Manutenção de Material Eletrotécnico	Nominal

		TMMT - Técnicos de Manutenção de Material Terrestre TOCAR - Técnicos de Operações de Circulação Aérea e Radar de Tráfego TOCC - Técnicos de Operações de Comunicações e Criptografia TODCI - Técnicos de Operações Detecção e Conduta de Interceção TOMET- Técnicos de Operações de Meteorologia TPAA - Técnicos de Pessoal e Apoio Administrativo TS - Técnicos de Saúde	
	Posto	Posto do militar com as seguintes categorias: 2SAR – Segundo-Sargento 1SAR – Primeiro-Sargento SAJ – Sargento-Ajudante SCH – Sargento-Chefe SMOR – Sargento-Mor ALF - Alferes TEN - Tenente CAP - Capitão MAJ - Major TCOR – Tenente-Coronel COR - Coronel	Ordinal
	Tmp_Posto	Tempo de permanência no posto até 31DEZ2023 ou à data de abate aos QP, em anos	Quantitativa
	Tmp_Serv_Efetivo	Tempo de serviço efetivo desde a incorporação até 31DEZ2023 ou à data de abate aos QP, em anos	Quantitativa
	Tmp_QP	Tempo de permanência nos quadros permanentes até 31DEZ2023 ou à data de abate aos QP, em anos	Quantitativa
	N_Louvres	Número de louvores atribuídos desde o ingresso nos QP	Quantitativa
	N_Condecorações	Número de condecorações atribuídas desde o ingresso nos QP	Quantitativa
	Nota_FAV	Nota global da última avaliação de mérito do militar, expressa numa escala de 1 a 5, denominada de Ficha de Avaliação Individual (FAV)	Quantitativa
	N_Colocações	Número de Unidades de Colocação desde o ingresso nos QP	Quantitativa
	Tmp_colocação	Tempo de permanência na Unidade até 31DEZ2023 ou à data de abate aos QP, em anos	Quantitativa
	Dist_UEO_RE S	Distância linear entre a Unidade de Colocação e a Residência. Para os militares colocados ou a residir nas ilhas e colocados no estrangeiro foi atribuída uma distância padrão de 500 km	Quantitativa
	N_Funções	Número de funções desempenhadas desde o ingresso nos QP	Quantitativa
	N_Formações	Número de formações obtidas desde o ingresso nos QP	Quantitativa
	Nota_ingresso	Nota final do curso de ingresso nos QP	Quantitativa

A tabela 1 apresenta as variáveis que foram selecionadas para este estudo. Pode-se verificar que existem 6 variáveis sociodemográficas e 16 estruturais. A variável **Abate** será a variável dependente do modelo de aprendizagem para prever se um determinado militar irá pedir abate aos QP.

Importa salientar que os dados têm como data de referência o dia 31 de dezembro de 2023, ou no caso dos militares que pediram abate aos QP a data efetiva da sua saída.

3. Apresentação e discussão dos resultados

Neste capítulo, serão apresentados os resultados obtidos, com o objetivo de responder à pergunta inicial e aos objetivos específicos.

Para o efeito, será realizada, inicialmente, uma análise exploratória dos dados, tendo por base os valores da variável dependente: abate ou não abate.

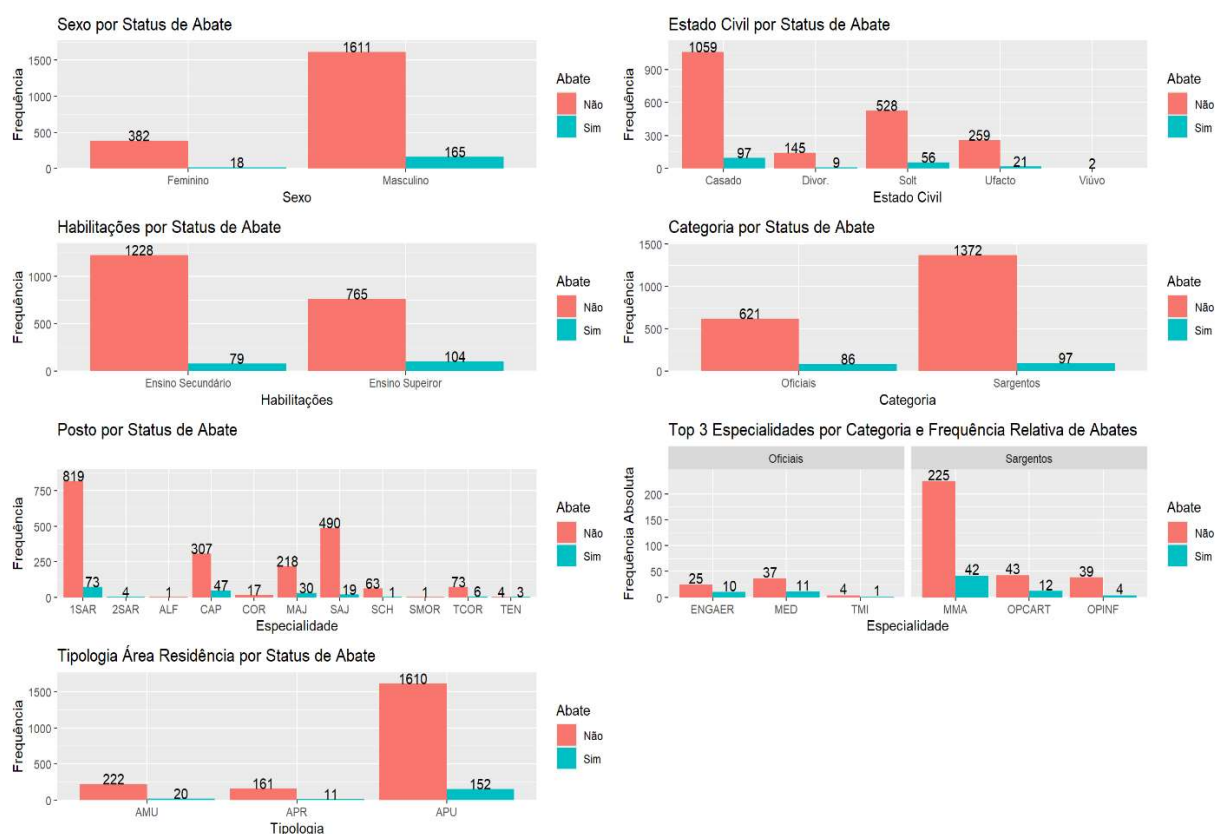
De seguida, procurar-se-á encontrar o melhor modelo de aprendizagem supervisionada, de entre os definidos no capítulo anterior.

Importa frisar que para a análise dos dados irá ser realizada em linguagem  utilizando para o efeito o software estatístico *R Studio*.

3.1 Análise exploratória dos dados

Tal como referido anteriormente, antes de se apresentar os resultados da aplicação dos métodos de aprendizagem supervisionada, é importante caracterizar a amostra em estudo, fazendo a divisão pelas dimensões que irão permitir classificar os dados.

Figura 4 – Distribuição das variáveis qualitativas



Relativamente ao sexo, a categoria com maior percentagem de abates foi a masculina, com 9,3%, enquanto no grupo feminino essa percentagem foi apenas 4,5%.

No que diz respeito ao estado civil, as categorias com maior número de abates foram os solteiros (9,6%) e os casados (8,4%).

Observa-se que a situação de abate é mais elevada entre os que possuem o Ensino Superior (12,0%), em comparação com os que possuem apenas o Ensino Secundário (6,0%). Além disso, verifica-se uma maior percentagem de abates entre os Oficiais (12,2%) face aos Sargentos (6,6%).

Quanto ao posto, as categorias que mais se destacam são os Capitães (CAP) com 13,3%, os Majores (MAJ) com 12,1% e os 1.º Sargentos (1SAR) com 8,2%.

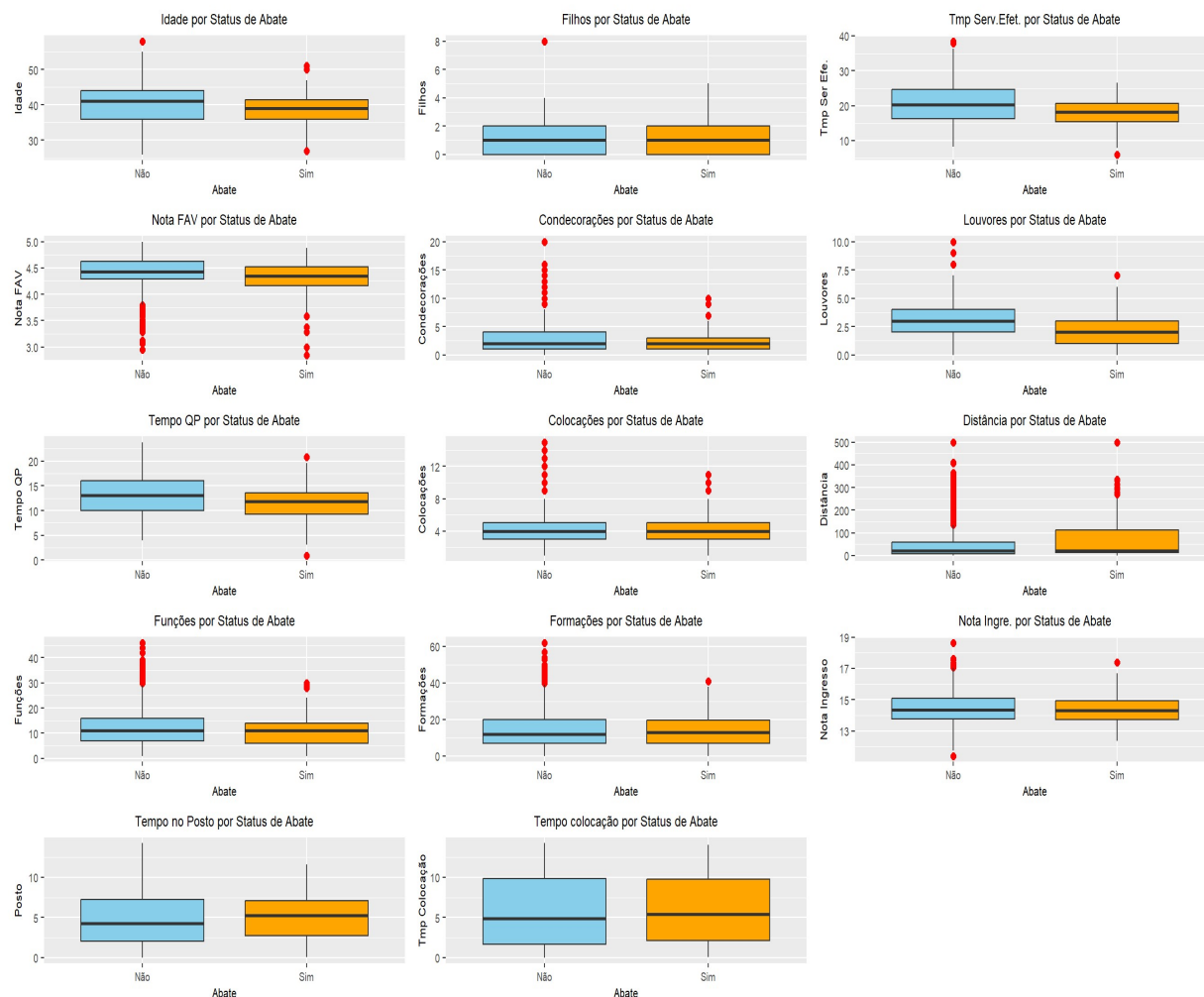
Nas especialidades dos Oficiais, destacam-se os Engenheiros Aeronáuticos (ENGAER) com 28,6% e os Médicos (MED) com 22,9%. Entre os Sargentos, as especialidades mais relevantes são os Operadores de Circulação Aérea e Radaristas de Tráfego (OPCART) com 21,8% e os Mecânicos de Manutenção Aeronáutica (MMA) com 15,6%.

Por fim, no que se refere à Tipologia da Área de Residência, existe uma maior percentagem de abates nas categorias de Áreas Predominantemente Urbanas (APU) com 8,6% e Áreas Moderadamente Urbanas (AMU) com 8,3%.

Tabela 2 – Estatísticas das variáveis quantitativas

Variáveis	Abate	Média	Mediana	Moda	D.P	Min.	Máx.	Q1	Q3
Número de Filhos	Não	1,15	1,00	1,00	0,92	0,00	8,00	0,00	2,00
	Sim	0,99	1,00	0,00	0,95	0,00	5,00	0,00	2,00
Idade	Não	40,20	41,00	42,00	5,36	26,00	58,00	36,00	44,00
	Sim	39,14	39,00	41,00	4,20	27,00	51,00	36,00	41,50
Tempo Serviço Efetivo	Não	20,42	20,29	19,55	5,55	8,16	38,48	16,17	24,61
	Sim	17,92	18,17	21,36	3,96	6,01	26,56	15,32	20,59
Tempo QP	Não	13,07	13,09	14,08	4,56	4,03	23,84	10,05	16,09
	Sim	11,45	11,83	11,43	3,49	0,96	20,78	9,29	13,53
Número Louvres	Não	2,84	3,00	2,00	1,78	0,00	10,00	2,00	4,00
	Sim	2,48	2,00	2,00	1,49	0,00	7,00	1,00	3,00
Número Condecorações	Não	2,72	2,00	2,00	2,11	0,00	20,00	1,00	4,00
	Sim	2,45	2,00	2,00	1,59	0,00	10,00	1,00	3,00
Nota ficha Avaliação	Não	4,43	4,43	4,47	0,25	2,95	5,00	4,29	4,62
	Sim	4,32	4,35	4,00	0,30	2,85	4,88	4,16	4,52
Número Colocações	Não	4,10	4,00	3,00	2,11	1,00	15,00	3,00	5,00
	Sim	3,80	4,00	4,00	1,85	1,00	11,00	3,00	5,00
Tempo Colocação	Não	5,95	4,85	14,25	4,76	0,01	14,25	1,66	9,81
	Sim	5,90	5,37	10,00	4,13	0,01	14,07	2,11	9,73
Distância UEO-RES	Não	70,37	22,33	500,00	121,15	0,58	500,00	8,76	59,89
	Sim	79,90	21,86	11,95	116,71	0,66	500,00	11,95	111,92
Número Funções	Não	12,44	11,00	6,00	7,50	1,00	46,00	7,00	16,00
	Sim	10,89	11,00	11,00	6,17	1,00	30,00	6,00	14,00
Número Formações	Não	14,26	12,00	7,00	8,88	0,00	62,00	7,00	20,00
	Sim	13,93	13,00	4,00	8,41	0,00	41,00	7,00	19,50
Nota Ingresso	Não	14,43	14,35	14,11	0,98	11,41	18,65	13,75	15,08
	Sim	14,32	14,33	15,20	0,92	12,37	17,39	13,74	14,95
Tempo no Posto	Não	4,76	4,25	1,25	3,34	0,01	14,25	2,00	7,25
	Sim	5,08	5,27	6,00	2,99	0,03	11,60	2,70	7,10

Figura 5 – Distribuição das variáveis quantitativas



Analisando as variáveis quantitativas, pela análise conjunta da tabela 2 e figura 5, observa-se que em relação às variáveis número de filhos, número de condecorações, número de colocações, número de funções, número de formações e nota de ingresso nos QP, não existe uma grande diferença visível ao nível da distribuição dos dois grupos (abate vs. não abate). Por outro lado, ao observar as variáveis idade, tempo de serviço efetivo, tempo nos QP, tempo no posto, nota da ficha de avaliação, distância entre a unidade de colocação e a residência e número de louvores, verifica-se que existem diferenças visíveis entre os dois grupos, sendo que os militares que pediram abate aos QP apresentam menor idade, menor tempo de serviço efetivo, menor tempo nos QP, menor nota na ficha de avaliação, menor número de louvores e um maior tempo no posto e maior distância entre as unidade de colocação e a residência. Observando ainda a figura 5, é possível constatar a presença de *outliers* em algumas variáveis. A tabela 3 apresenta o número de valores considerados *outliers* nos dados das diversas variáveis analisadas. Verifica-se que a maioria destes valores está concentrado em

3 variáveis: distância entre a unidade e a residência (315); número de colocações (70) e número de funções (63). Importa ainda referir que a maioria destes valores corresponde a militares que não pediram abate aos QP.

Tabela 3 – Contabilização de *outliers* por variáveis e status de Abates

Variáveis	Não	Sim	Total
Número de Filhos	1	0	1
Idade	1	3	4
Tempo Serviço Efetivo	2	1	3
Tempo QP	0	2	2
Número Louvores	29	2	31
Número Condecorações	45	4	49
Nota ficha Avaliação	32	5	37
Número Colocações	65	5	70
Tempo Colocação	0	0	0
Distância UEO-RES	295	20	315
Número Funções	58	5	63
Número Formações	31	1	32
Nota Ingresso	10	1	11
Tempo no Posto	0	0	0
Total	568	49	617

Perante estes resultados, é fundamental refletir sobre o tratamento a ser dado aos *outliers*.

De acordo com Maroco (2003) e Gelman & Hill (2006) os valores extremos não devem ser eliminados à partida, pois podem revelar um padrão importante da amostra e que deve ser estudado.

Na mesma medida, Taleb (2011) compara os *outliers* aos “cisnes negros”, uma vez que estes representam eventos que não se espera que aconteçam. Assim, para o autor, quando não se inclui *outliers*, não se está a considerar acontecimentos passíveis de ocorrer. Isto significa que os *outliers* são menos prováveis de ocorrer e que não devem ser ignorados.

Assim, optou-se por incluir os *outliers* no estudo. Para esta decisão pesou também o facto destes valores corresponderem a 537 militares² que teriam de ser excluídos da investigação, representando aproximadamente 25% da amostra.

Resumindo, o perfil do militar que pede o abate aos QP, com base nas variáveis estudadas e consideradas mais discriminatória, é o seguinte:

Militar do sexo masculino, com uma média de 39 anos de idade, solteiro, tem o ensino superior e reside numa área predominantemente urbana, aproximadamente a 80km, em média, da unidade de colocação. Em relação à carreira militar, é oficial com o posto de Capitão, com uma média de 18 anos de serviço efetivo, 11 anos de QP e 5 anos de posto, tendo recebido,

² Dos 617 militares que apresentam *outliers* nas diferentes variáveis, verificou-se que existiam militares repetidos, pelo que o valor final é de 537.

em média, 2 louvores durante a sua permanência nos QP e a nota da sua última avaliação foi de 4,32.

3.2 Seleção do modelo

Tal como foi referido no capítulo anterior, foram considerados cinco métodos, RL, NB, LDA, QDA e AD, com o objetivo de classificar a situação dos militares relativamente à situação de abate aos QP.

A matriz de confusão será utilizada para avaliar o desempenho dos modelos de classificação em estudo, comparando os resultados previstos pelos modelos com os valores reais dos dados (Chawla et al., 2020; James et al., 2023; Lantz, 2013).

Para uma classificação binária (“sim” ou “não”), como se pode verificar na figura 6, a matriz de confusão organiza os resultados de previsão em quatro categorias:

- **Verdadeiro Negativo (VN)**: Casos em que o modelo previu "não" e a classe real também era "não";
- **Falso Negativo (FN)**: Casos em que o modelo previu "não", mas a classe real era "sim";
- **Verdadeiro Positivo (VP)**: Casos em que o modelo previu "sim" e a classe real também era "sim";
- **Falso Positivo (FP)**: Casos em que o modelo previu "sim", mas a classe real era "não".

A matriz de confusão permite aferir o número de erros produzidos por um determinado modelo de classificação, sendo que os valores na diagonal principal, assinalados a verde, indicam o número de previsões corretas e os da diagonal secundária, a vermelho, representam o número de previsões incorretas (Chawla et al., 2020; James et al., 2023; Lantz, 2013).

Figura 6 – Matriz de confusão

		PREVISTO	
		Não	Sim
REAL	Não	VN – Verdadeiro Negativo	FP – Falso positivo
	Sim	FN – Falso Negativo	VP – Verdadeiro Positivo

Fonte: Adaptado de Lantz (2013) e James et al. (2023)

A partir da matriz de confusão pode-se calcular um conjunto de métricas que permitem validar a performance do modelo, sendo a mais comum a taxa de exatidão que indica a taxa de

sucesso na identificação correta dos valores verdadeiros (Chawla et al., 2020; James et al., 2023; Lantz, 2013).

A matriz de confusão permite as seguintes métricas:

- **Acurácia (Taxa de Exatidão):** Proporção total de previsões corretas;
- **Precisão:** Proporção de verdadeiros positivos entre todos os positivos previstos. Indica quantos dos casos classificados como positivos são realmente positivos. **Alta precisão** significa que a maioria dos casos que o modelo classifica como positivos são realmente positivos, minimizando falsos positivos. Isso é especialmente desejável em contextos onde os falsos positivos têm um custo elevado (ex.: diagnósticos médicos, onde é importante evitar falsos alarmes). **Baixa precisão** indica que o modelo está a fazer muitas previsões positivas incorretas. A precisão, portanto, é uma métrica essencial quando o objetivo é ter confiança nos casos classificados como positivos;
- **Recall (Sensibilidade):** Proporção de verdadeiros positivos entre todos os positivos reais. Mede a capacidade do modelo em identificar corretamente os casos positivos entre todos os casos que são realmente positivos. **Alta sensibilidade** significa que o modelo é eficaz em capturar a maioria dos casos positivos reais, minimizando falsos negativos. Isso é particularmente importante em contextos onde é crítico identificar todos os casos positivos, como em diagnósticos de doenças graves (onde se quer evitar deixar de diagnosticar casos positivos). **Baixa sensibilidade** indica que o modelo tende a classificar incorretamente muitos casos positivos como negativos.
- **Especificidade:** Proporção de verdadeiros negativos entre todos os negativos reais. Mede a capacidade do modelo em identificar corretamente os casos negativos entre todos os casos que são realmente negativos. **Alta especificidade** significa que o modelo é eficaz em evitar falsos positivos, classificando corretamente os casos negativos. Isto é desejável em contextos onde é importante minimizar falsos positivos (ex.: diagnósticos para uma doença rara, onde é melhor evitar diagnósticos incorretos para quem não tem a condição). **Baixa especificidade** indica que o modelo tende a classificar incorretamente casos negativos como positivos.
- **F1 Score:** É a média harmónica entre precisão e sensibilidade, o que significa que penaliza mais fortemente valores baixos em qualquer uma dessas métricas. Como média harmónica, o F1-score é mais sensível a valores baixos e, portanto, apenas assume um valor elevado se ambas as métricas (precisão e sensibilidade) forem altas. Um F1-score alto significa que o modelo tem boa precisão e sensibilidade, ou seja, consegue identificar corretamente a maioria dos casos positivos sem gerar muitos falsos positivos. Um F1-score baixo indica que o modelo tem dificuldade em manter alta tanto a precisão quanto a sensibilidade, resultando num número considerável de falsos positivos ou falsos negativos.

Para Lantz (2013), criar um algoritmo de classificação envolve muitas vezes o equilíbrio entre escolher ser demasiado conservador ou demasiado agressivo na tomada de decisão. Esta escolha reflete-se na forma como o modelo lida com os diferentes tipos de erro, como os falsos positivos e os falsos negativos. Ser conservador implica que o modelo pode ser mais cauteloso nas suas previsões, tendendo a reduzir os falsos positivos. No entanto, esta abordagem pode aumentar o número de falsos negativos, pois o modelo hesita em classificar um caso como positivo. Ser agressivo implica que o modelo está mais inclinado a classificar casos como positivos, aumentando a sensibilidade (capacidade de identificar verdadeiros positivos). No entanto, isso pode levar a um aumento de falsos positivos, pois o modelo assume mais riscos na classificação.

Tabela 4 – Métricas de performance do modelo

Exatidão	$\frac{VP + VN}{VP + VN + FP + FN}$
Precisão	$\frac{VP}{VP + FP}$
Sensibilidade (<i>recall</i>)	$\frac{VP}{VP + FN}$
Especificidade	$\frac{VN}{VN + FP}$
F1-Score	$2 \times \frac{\text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}$

Fonte: Adaptado de Lantz (2013)

A tabela 4 apresenta as fórmulas utilizadas para o cálculo destas cinco métricas, sendo importante referir que todos se expressam no intervalo entre 0 e 1, sendo desejável que todos os indicadores se aproximem de 1 para conferir robustez ao modelo.

Ao aplicar os cinco métodos aos nossos dados, foi possível obter uma matriz de confusão para cada um deles (Apêndice 1), e calcular as cinco métricas referidos na tabela 4.

Na tabela 5 podem-se observar os resultados destes cinco valores para cada um dos métodos de classificação utilizados, verificando-se que todos apresentam valores de exatidão superiores a 80%, o que indica que conseguem prever com alguma exatidão quem permanece na organização e quem pede abate aos QP da FA.

No entanto, quando se fala em precisão, ou seja, a proporção de verdadeiros casos serem abate de entre todos os que o modelo considerou abate, verifica-se que apenas o RL (64%), o AD (53%) e o LDA (57%) apresentam taxas acima dos 50%.

No que diz respeito à sensibilidade verifica-se que todos os métodos não conseguem classificar corretamente todos os casos positivos, ou seja, os militares que pediram abate aos QP, com a eficácia desejada. O método NB apresenta o valor mais elevado com 44% e o AD e o QLA apresentam o valor mais baixo com 15%.

Em relação à especificidade constata-se que todos os métodos conseguem atingir uma taxa superior a 85% em classificar corretamente todas as observações negativas, ou seja, os militares que não pediram abate aos QP.

Por último, ao observar o resultado para o *F1-Score*, verifica-se que o algoritmo que obtém o melhor valor é o RL (43%), seguido do LDA (39%). Por outro lado, o valor mais baixo é conseguido pelo QDA com 19%. Estes resultados são fortemente influenciados pelos níveis baixos de sensibilidade presentes nestes algoritmos.

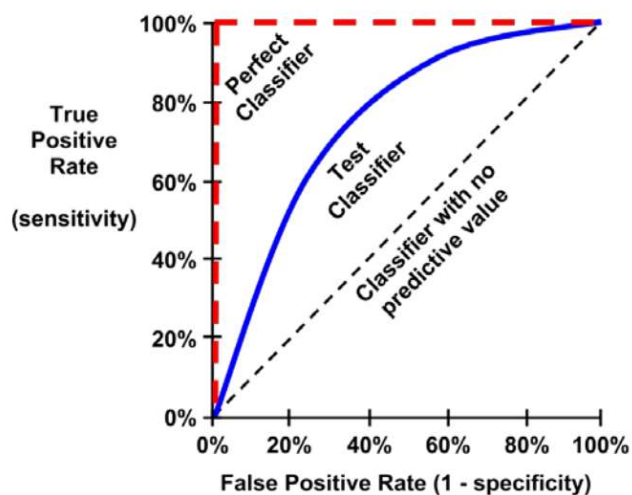
Tabela 5 – Resultados matriz de confusão dos métodos

Método	Exatidão	Precisão	Sensibilidade	Especificidade	F1-score
RL	0,93	0,64	0,33	0,98	0,44
AD	0,92	0,53	0,15	0,99	0,23
NB	0,82	0,22	0,44	0,86	0,29
LDA	0,92	0,57	0,29	0,98	0,38
QLA	0,90	0,29	0,15	0,97	0,20

No entanto, estes cinco indicadores não são a única maneira de validar a eficácia de um modelo. De acordo com James et al. (2023), Lantz (2013), Marvin et al. (2021) e Yedida et al. (2024), a curva ROC (*Receiver Operating Characteristic*) é uma representação gráfica útil para avaliar a eficácia de um modelo de *machine learning*, particularmente nos modelos de classificação binária. A curva ROC é especialmente valiosa quando é necessário encontrar um equilíbrio adequado entre sensibilidade e especificidade. Essa necessidade de equilíbrio é crucial em aplicações onde os custos de erros de classificação não são uniformes — por exemplo, em diagnósticos médicos, onde uma alta sensibilidade pode ser preferida para evitar falsos negativos perigosos, mas uma especificidade alta também é importante para evitar falsos positivos, e, conseqüentemente, falsos alarmes que podem levar a exames e tratamentos adicionais. Assim, a curva ROC fornece uma representação visual intuitiva para

comparar o desempenho do modelo e ajustar os limiares de decisão de acordo com os requisitos específicos do problema.

Figura 7 – Representação de uma curva ROC



Fonte: Lantz (2013)

A figura 7 representa uma curva ROC, indicando a taxa de verdadeiros positivos (sensibilidade) no eixo vertical e a taxa de falsos positivos (1-especificidade) no eixo horizontal, sendo que uma curva ROC ideal aproxima-se do canto superior esquerdo, indicando uma alta taxa de verdadeiros positivos e uma baixa taxa de falsos positivos.

Um classificador perfeito - Curva ROC Perfeita - teria uma taxa de verdadeiros positivos de 100% e uma taxa de falsos positivos de 0%, resultando numa curva que começaria no ponto (0,0) para (0,100) e depois para (100,100), formando um canto superior esquerdo. Essa situação indicaria que o modelo conseguiria discriminar perfeitamente as classes, sem cometer nenhum erro.

Um classificador aleatório - Curva ROC Aleatória - não possui capacidade de discriminar entre as classes positiva e negativa, agindo essencialmente como um modelo que efetua previsões de forma totalmente aleatória. Graficamente, essa curva é uma linha diagonal que vai do ponto (0,0) ao ponto (100,100), o que significa que, à medida que a taxa de verdadeiros positivos aumenta, a taxa de falsos positivos aumenta na mesma proporção. Quando a Área Sob Essa Curva (AUC, do inglês *Area Under the Curve*) é de 0,5, indica que o modelo não tem melhor desempenho do que uma simples escolha aleatória.

Desta forma, o desempenho global de um modelo de classificação pode ser medido através da AUC, sendo que um valor de 1 representa um modelo perfeito e um valor de 0,5 representa um modelo aleatório (Lantz, 2013). Quanto maior o AUC, melhor o desempenho geral do modelo (James et al., 2023).

De acordo com Hosmer et al. (2013) não existe um número mágico que nos permita aferir uma boa discriminação pela AUC, apenas algumas orientações gerais que podem ser estabelecidas, conforme exposto na Tabela 6.

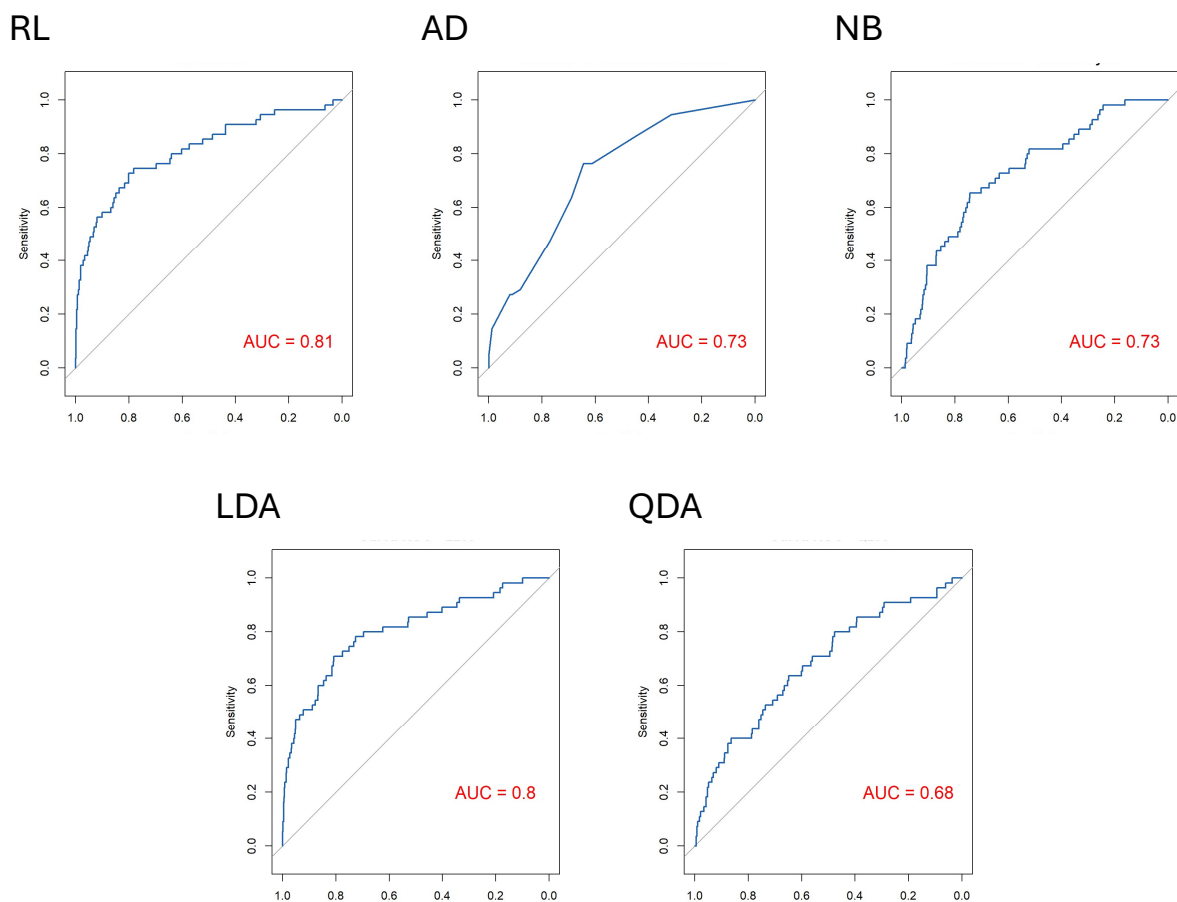
Tabela 6 – Interpretação da Pontuação AUC

Pontuação	Interpretação
0,9 - 1	Discriminação Extraordinária
0,8 – 0,9	Discriminação Excelente
0,7 – 0,8	Discriminação aceitável
0,5 – 0,7	Discriminação pobre
< 0,5	Sem discriminação

Fonte: adaptado de (Hosmer et al., 2013)

Observando a figura 8, pode-se verificar que todos os modelos estudados apresentam valores AUC discriminatórios, ou seja, superiores a 0,6. No entanto, temos de destacar pela positiva a RL e a LDA que apresentam valores de AUC de 0,81 e 0,80, respectivamente, o que nos indica alguma capacidade preditiva destes modelos.

Figura 8 – ROC Curves dos métodos



Perante estes resultados, torna-se então possível selecionar o modelo de classificação que melhor se adequa aos dados. Assim, tendo em consideração os valores das métricas fornecidos pela matriz de confusão e os resultados obtidos na curva ROC, considera-se que todos os modelos apresentam valores favoráveis para a continuação da análise.

No entanto, optou-se por selecionar a RL por ser o método que melhor prevê os abates aos QP da FA, uma vez que é o que apresenta melhores valores nestas duas análises.

3.2.2 Aperfeiçoamento do modelo

Após optar pelo método da RL, o próximo passo é aperfeiçoá-lo, priorizando a eficácia, procurando também torná-lo mais simples e fácil de entender. Para o efeito, existem procedimentos que permitem selecionar as variáveis que possuem as relações mais fortes com a variável dependente, sendo essas as que mais contribuem para a previsão (Maroco, 2021).

Segundo James et al. (2023) e Maroco, (2021), um dos métodos para a seleção de variáveis é *Stepwise*, que pode ser aplicado através de três abordagens: seleção *Forward*, seleção *Backward* e seleção *Stepwise* ou Híbrida.

A seleção *Forward* inicia a análise sem variáveis preditoras e vai adicionando-as uma a uma, selecionando em cada etapa a variável que faz melhorar o ajustamento do modelo aos dados, a partir de um determinado critério (James et al., 2023). Em cada passo, a variável que, ao ser incluída, minimiza o erro ou melhora a métrica escolhida (como por exemplo o *Akaike Information Criterion*, AIC) é adicionada ao modelo. O processo continua até que não seja possível adicionar nenhuma variável que traga uma melhoria significativa, conforme o critério estabelecido. Esta abordagem é útil porque constrói o modelo de forma incremental, facilitando a identificação das variáveis que realmente têm impacto na previsão, sem sobrecarregar o modelo com variáveis irrelevantes. No entanto, ela pode ser menos eficaz quando há forte correlação entre as variáveis, o que pode levar à seleção incorreta de variáveis importantes.

Por outro lado, a seleção *Backward* parte de um modelo que inclui todas as variáveis independentes, removendo, uma a uma, aquelas que se mostram menos significativas para o modelo (Maroco, 2021). Em cada etapa, a variável cuja remoção resulta na menor perda de qualidade do modelo é excluída. Essa decisão é feita com base num critério estatístico, como o AIC, ou através da significância estatística das variáveis (por exemplo, utilizando os valores *p-value*). O procedimento continua até que a exclusão de qualquer outra variável resulte numa deterioração considerável do desempenho do modelo. No entanto, este método pode não ser adequado em situações onde o número de variáveis independentes é muito elevado, ou quando existem fortes correlações entre elas.

A seleção *Stepwise* é um método híbrido, que combina as abordagens *Forward* e *Backward* para otimizar o processo de escolha das variáveis preditoras de forma mais dinâmica e flexível. Neste método, à medida que as variáveis vão sendo adicionadas ao modelo, as que já existem são novamente testadas, podendo ser removidas se a sua importância no modelo diminuir (James et al., 2023; Maroco, 2021). A seleção *Stepwise* permite um ajustamento mais preciso ao considerar simultaneamente a inclusão e a exclusão de variáveis. O método permite evitar a inclusão de variáveis irrelevantes e eliminar aquelas que não contribuem de forma significativa para o modelo. O método é útil para encontrar um modelo equilibrado e parcimonioso, especialmente em situações onde há um grande número de variáveis independentes ou quando existe a possibilidade de correlações significativas entre elas.

Para este estudo, optou-se pela utilização do método híbrido, recorrendo ao critério AIC e aos valores de *p-value*, uma vez que é o mais apropriado quando existem correlações significativas (apêndice 2) entre as variáveis independentes (Maroco, 2021) e combina a eficiência dos dois métodos *Forward* e *Backward* (James et al., 2023).

Tabela 7 – Resultados RL antes e depois da seleção *Stepwise*

Métrica		Antes <i>Stepwise</i>	Após <i>Stepwise</i>
Variáveis		21	12
Resíduos de Desvio	Mínimo	-1,83164	-1,88411
	1º Quartil	-0,36787	-0,37428
	Mediana	-0,18644	-0,18439
	3º Quartil	-0,06318	-0,06689
	Máximo	3,11936	3,04497
Desvio nulo		878,89 (1522 df)	878,89 (1522 df)
Desvio residual		600,34 (1451 df)	603,98 (1464 df)
AIC		744,34	721,98

Pela análise da Tabela 7, pode-se verificar que existiu um ajuste do modelo inicial (apêndice 3), após a aplicação do método *Stepwise*, resultando na eliminação de 9 variáveis independentes. Esse procedimento permitiu simplificar o modelo, removendo variáveis que não contribuíam significativamente para a previsão, de acordo com os critérios estatísticos utilizados no processo de seleção. Assim, o modelo final tornou-se mais parcimonioso, mantendo apenas as variáveis com maior relevância para a explicação da variável dependente.

Segundo Hosmer et al. (2013), uma das medidas mais utilizadas para comparar modelos é o AIC, onde geralmente é preferível utilizar valores mais baixos. Um valor baixo de AIC indica um modelo melhor, pois sugere um ajustamento mais eficiente do modelo aos dados, sem adicionar complexidade desnecessária. Também de referir que o AIC penaliza modelos que têm muitos parâmetros. Isso significa que, embora um modelo mais complexo possa ajustar-

se melhor aos dados, o AIC incorpora um ajuste que favorece modelos mais simples, desde que a qualidade do ajustamento não seja significativamente reduzida.

Assim, verifica-se a redução do valor de AIC de 744,34, do modelo inicial com as 21 variáveis, para 721,98 do modelo com 12 variáveis, sugerindo que este modelo é mais simples e que existe uma melhor relação entre a complexidade e a qualidade no ajustamento do modelo aos dados.

A análise dos resíduos, tendo em consideração que não existiram alteração significativas aos valores, sugere que o modelo melhorado se manteve estável em relação ao inicial.

Tabela 8 – Resultado das variáveis do modelo após a seleção *Stepwise*

Variável	Fatorial	Coefficiente	Erro est.	Estatística	p-value
(Constante)		-7,669	3,629	-2,113	0,035 *
Filhos		-0,199	0,131	-1,515	0,130
Idade		0,463	0,065	7,171	0,000 ***
Habilitações	Ensino Superior	1,324	0,413	3,204	0,001 **
Posto	2SAR	35,380	5107,000	0,007	0,994
	ALF	3,036	11280,000	0,000	1,000
	CAP	-1,824	1,120	-1,628	0,103
	COR	-20,110	2802,000	-0,007	0,994
	MAJ	-1,172	1,333	-0,879	0,379
	SAJ	0,146	0,784	0,187	0,852
	SCH	-0,431	1,631	-0,264	0,792
	SMOR	-15,530	10750,000	-0,001	0,999
	TCOR	-3,990	1,762	-2,265	0,024 *
	TEN	1,230	2,501	0,492	0,623
Tempo Posto		0,160	0,080	1,995	0,046 *
Nº Louvores		0,204	0,094	2,172	0,030 *
Nº Formações		-0,096	0,023	-4,171	0,000 ***
Sexo	Masculino	0,939	0,423	2,220	0,026 *
Tmp Serviço Efetivo		-0,411	0,072	-5,712	0,000 ***
Nota Ingresso		-0,211	0,135	-1,561	0,119
Especialidade	ADMAER	-15,080	1751,000	-0,009	0,993
	CMI	0,156	1,322	0,118	0,906
	ENGAED	2,753	1,631	1,688	0,091
	ENGAER	5,417	1,546	3,504	0,000 ***
	ENGEL	3,037	1,573	1,930	0,054
	JUR	-15,960	4791,000	-0,003	0,997
	MARME	2,020	0,964	2,096	0,036 *
	MED	3,489	1,551	2,249	0,025 *
	MELECA	0,297	0,980	0,304	0,761
	MELECT	-16,110	1625,000	-0,010	0,992
	MELIAV	2,348	0,969	2,424	0,015 *
	MMA	2,874	0,881	3,262	0,001 **
	MMT	-0,347	1,307	-0,266	0,791
	MUS	-1,779	1,342	-1,325	0,185
	NAV	1,363	1,773	0,769	0,442
	OPCART	2,696	1,004	2,685	0,007 **
	OPCOM	-1,056	1,293	-0,816	0,414
	OPINF	0,462	1,325	0,348	0,728
	OPMET	-16,500	2160,000	-0,008	0,994
	OPRDET	0,248	1,320	0,188	0,851
	OPSAS	0,376	1,333	0,282	0,778
	PA	0,693	1,028	0,674	0,500

	PILAV	4,240	1,489	2,848	0,004 **
	PSI	2,255	1,780	1,267	0,205
	SAS	-1,503	1,097	-1,371	0,171
	TABST	-16,790	2395,000	-0,007	0,994
	TINF	1,794	1,498	1,198	0,231
	TMAEQ	-16,830	3398,000	-0,005	0,996
	TMI	-16,800	5087,000	-0,003	0,997
	TMMA	2,140	1,515	1,413	0,158
	TMMEI	0,962	1,725	0,558	0,577
	TMMT	0,920	1,744	0,528	0,598
	TOCART	0,043	1,715	0,025	0,980
	TOCC	-16,850	3318,000	-0,005	0,996
	TODCI	-16,410	2005,000	-0,008	0,993
	TOMET	-16,490	3573,000	-0,005	0,996
	TPAA	-0,767	1,704	-0,450	0,653
	TS	-16,300	3349,000	-0,005	0,996
Nota FAV		-1,046	0,472	-2,217	0,027 *

p-value: '****' 0.001 '***' 0.01 '*' 0.05 '.' 0.1

A Tabela 8 apresenta os valores das doze variáveis remanescentes no modelo, após a aplicação do método *stepwise*, sendo possível analisar o impacto de cada uma delas na determinação da probabilidade de abate e definir quais as que são mais significativas para a explicação do abate aos QP. Para analisar a significância dessas variáveis no modelo, considera-se o *p-value* associado a cada uma delas. No âmbito deste estudo, o *p-value* de referência será igual a 0,05, o que significa que uma variável será considerada estatisticamente significativa se o seu *p-value* for menor ou igual a esse limiar, significando que tem um impacto relevante e é considerada importante para a explicação do fenómeno em análise. Dessa forma, as variáveis com *p-values* menores que 0,05 são aquelas que efetivamente contribuem para o modelo, enquanto as variáveis com *p-values* maiores que 0,05 podem não ser tão relevantes, podendo ser eliminadas do modelo.

Analisando a variável *Filhos*, observa-se um coeficiente estimado de -0,199, o que sugere que, para cada filho adicional, a probabilidade de saída da FA tende a diminuir. Contudo, como o valor do *p-value* é superior a 0,05, este efeito não é estatisticamente significativo, impossibilitando uma conclusão fiável sobre a influência do número de filhos na probabilidade de saída da FA.

Quanto à variável *Idade*, observa-se um impacto significativo sobre a variável *Abate*, com um coeficiente positivo de 0,463 e um *p-value* muito baixo (0,001), indicando uma forte influência na saída da organização. Querendo isto dizer que, cada aumento unitário na idade aumenta a probabilidade de saída.

Com um coeficiente estimado de 1,324 e um *p-value* inferior a 0,001, a variável *Habilitações* indica que militares com ensino superior têm uma maior probabilidade de pedir abate aos QP, em comparação com aqueles sem ensino superior, e que esse efeito é estatisticamente significativo.

Em relação à variável *Posto* pode-se concluir que apenas o posto de TCOR apresenta um coeficiente (-3,990) estatisticamente significativo ($p\text{-value} < 0,05$), sugerindo que este tem uma menor probabilidade de saída face à categoria de referência (1SAR). Os restantes postos, apesar de alguns apresentarem coeficientes elevados, não são estatisticamente significativos, uma vez que os valores de significância são superiores a 0,05.

Para a variável *Tempo de Permanência no Posto*, o coeficiente de 0,16 é estatisticamente significativo ($p\text{-value} < 0,05$), indicando que cada unidade adicional de tempo no posto aumenta a probabilidade de saída da FA. Assim, conclui-se que esta variável exerce um impacto positivo na probabilidade de pedido de abate aos QP.

Para a variável *Número de Louvores*, observa-se um coeficiente positivo de 0,204, estatisticamente significativo ($p\text{-value} < 0,05$), indicando que militares com mais louvores têm uma maior probabilidade de sair da organização. É importante destacar que este resultado não implica que militares sem louvores não saiam, mas sugere que o reconhecimento do talento através de louvores pode torná-los mais suscetíveis a convites de outras organizações. No que diz respeito à variável *Número de Formações*, verifica-se que existe um impacto negativo na probabilidade de saída, uma vez que o coeficiente é de -0,096 e significativo a nível estatístico ($p\text{-value} < 0,001$), sugerindo que participar em mais formações está associado a uma menor probabilidade de sair da FA.

Para a variável *Sexo*, verifica-se que os homens têm uma maior probabilidade em sair da FA, quando comparados com as mulheres, com um coeficiente estatisticamente significativo ($p\text{-value} < 0,05$) de 0,939, ou seja, ser homem está associado a uma maior probabilidade de saída.

Relativamente à variável *Tempo de Serviço Efetivo* observa-se um coeficiente estimado de -0,411, sugerindo que, por cada unidade de tempo de serviço efetivo, a probabilidade de saída da organização diminui, ou seja, quanto mais tempo o militar estiver na FA, menor é a probabilidade de ele sair. O $p\text{-value}$ é inferior a 0,001 o que sugere que tem um impacto estatisticamente significativo e forte na redução da probabilidade de pedido de abate aos QP. Analisando a variável *Nota de Ingresso*, constata-se que apesar do coeficiente estimado ser de -0,211, sugerindo que por cada ponto adicional na nota de ingresso a probabilidade de saída da organização diminui, verifica-se que esta relação não é estatisticamente significativa ($p\text{-value} > 0,05$), pelo que não se pode afirmar com confiança que a nota de ingresso tem um efeito na probabilidade de saída.

Relativamente à variável *Especialidade* verifica-se que as categorias que apresentam coeficientes estatisticamente significativos são ENGAER (coef. 5,417; $p\text{-value} < 0,001$), Mecânicos de Armamento e Equipamento (MARME) (coef. 2,020; $p\text{-value} < 0,05$), MED (coef. 3,489; $p\text{-value} < 0,05$), Mecânicos de Eletricidade e Instrumentos de Avião (MELIAV) (coef. 2,348; $p\text{-value} < 0,05$), MMA (coef. 2,874; $p\text{-value} < 0,01$), OPCART (coef. 2,696; $p\text{-value}$

<0,01) e Piloto-Aviador (PILAV) (coef. 4,240; *p-value* <0,01), o que indica que estas especialidade estão associadas a uma maior probabilidade de abate aos QP face à categoria de referência que foi Abastecimento (ABST).

As restantes especialidades não apresentam evidências de terem um impacto relevante na probabilidade de saída, de acordo com os *p-value* analisados.

Por último, para a variável *Nota FAV* o coeficiente estimado tem um valor negativo, -1,046, com um *p-value* <0,05, sugerindo a existência de uma relação estatisticamente significativa, em que uma maior nota na FAV provoca uma menor probabilidade de pedido de abate aos QP.

Em suma, após a aplicação do método *Stepwise*, das 12 variáveis remanescentes, somente duas — *Número de Filhos* e *Nota de Ingresso* — não apresentaram valores estatisticamente significativos. Isso indica que não existem evidências fortes que sustentem que essas variáveis estejam relacionadas com a probabilidade de saída da FA, enquanto as demais demonstraram uma associação significativa com o abate aos QP.

Diante destes resultados, conclui-se que o modelo ideal não deve incluir essas duas variáveis, razão pela qual foram excluídas da versão final. Desta forma, foi construído um modelo denominado de RL 1.0 que contempla as 10 variáveis consideradas significativas para a explicação do abate aos QP.

Figura 9 – Matriz confusão RL 1.0

		PREVISTO	
		Não	Sim
REAL	Não	591	7
	Sim	34	21

Tabela 9 – Resultados matriz de confusão RL original vs. RL 1.0

Algoritmo	RL	RL 1.0
Exatidão	0,93	0,94
Precisão	0,64	0,75
Sensibilidade	0,33	0,38
Especificidade	0,98	0,99
F1-Score	0,43	0,50

A Tabela 9 compara o desempenho dos dois modelos de RL, o original e melhorado (RL 1.0), após a aplicação do método *Stepwise* e análise da significância. Foram aplicadas as métricas utilizadas anteriormente, exatidão, precisão, sensibilidade, especificidade e F1-Score, extraídas da análise da Figura 9, que contém os resultados da matriz de confusão.

Assim, pode-se constatar que em termos de exatidão, o modelo RL 1.0 apresentou uma ligeira melhoria no desempenho, de 93% para 94%, indicando que apesar de ambos os modelos terem uma alta capacidade de classificar corretamente os casos, a redução do número de variáveis resultou numa leve melhoria nesse indicador.

Ao nível da precisão, verifica-se que o modelo melhorado aumentou significativamente de 64% para 75%, indicando que se verificou um aumento na probabilidade deste modelo classificar corretamente um caso de abate, em relação ao total de previsões classificadas como abate.

Observando a sensibilidade, ou seja, a capacidade do modelo em classificar corretamente um abate aos QP (isto é, identificar corretamente os casos positivos), verifica-se uma melhoria no modelo RL 1.0 (33% para 38%). No entanto, apesar desta melhoria, o valor ainda é relativamente baixo, indicando que ambos os modelos têm dificuldade em identificar corretamente os casos positivos. Estes resultados vêm comprovar o que já foi defendido por alguns autores, como Chiavenato (2014), Halim et al. (2020), Hom et al. (2017) e Lee & Mitchell (1994), sobre a complexidade da decisão em sair de uma organização quando existem diversos fatores que podem influenciar esta decisão.

Relativamente à especificidade, observa-se que ambos os modelos são muito eficazes em identificar corretamente os casos negativos, ou seja, os militares que não saem da organização, apesar de existir uma ligeira melhoria entre os modelos (98% para 99%).

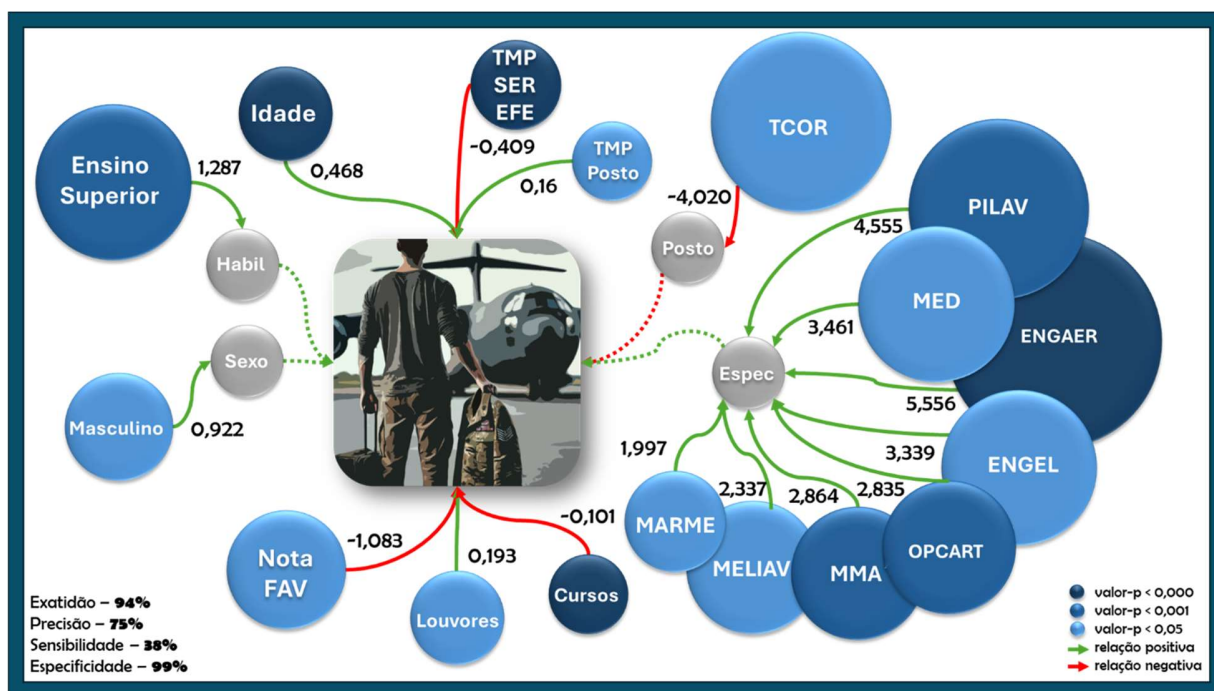
Por último, existe uma melhoria do *F1-Score* no RL 1.0, que passou de 44% para 50%, sendo esta a medida que combina os valores de precisão e sensibilidade. Apesar de se verificar uma

melhoria, o valor indica que o modelo RL 1.0 pode não estar a conseguir captar a complexidade em identificar os casos positivos.

Em suma, a comparação entre estes dois modelos de RL revela melhorias em várias métricas no modelo RL 1.0, nomeadamente na exatidão, precisão, especificidade e F1-Score. No entanto, a sensibilidade permanece baixa, indicando a necessidade de análises adicionais para melhorar a taxa de deteção dos casos positivos.

Consequentemente, com base nos resultados do modelo RL 1.0 (Apêndice 4), é possível representá-lo graficamente para visualizar as conexões e identificar as variáveis que mais contribuem para o pedido de abate.

Figura 10 – Representação das variáveis significativas do modelo RL 1.0



A figura 10 apresenta as variáveis mais significativas do modelo RL 1.0 onde se observa o comportamento das 10 variáveis remanescentes depois de se aperfeiçoar o modelo. Assim, constata-se que as variáveis *Idade*, *Habilidades Literárias*, *Sexo*, *Nota FAV*, *Nº Louvores*, *Nº Formações*, *Especialidade*, *Posto*, *Tempo no Posto* e *Tempo de Serviço Efetivo* são aquelas que conseguem prever melhor o comportamento do militar quando este decide pedir abate aos QP. É possível observar pelo **tamanho do círculo** quais as variáveis que têm um maior impacto na variável *Abate*. Os círculos maiores indicam variáveis com um impacto mais significativo. **As cores dos círculos** refletem a força da relação: cores mais escuras representam associações mais fortes, enquanto cores mais claras indicam associações mais fracas com a variável dependente. Além disso, as **cores das setas** indicam a direção do impacto: o verde, associado a um impacto positivo, significa que a variável aumenta a probabilidade de abate, enquanto o vermelho, associada a um impacto negativo, indica que a variável diminui essa probabilidade. Verifica-se também que em relação à análise efetuada anteriormente, quando se aplicou o método *Stepwise* (Tabela 8), o impacto positivo, ou negativo, das variáveis independentes na variável *Abate* mantém-se inalterado, bem como nos valores dos coeficientes que apresentaram diferenças mínimas em relação aos valores do *Stepwise*.

Por outro lado, é importante salientar que, na variável *Especialidade*, a categoria Engenheiros Eletrotécnicos (ENGEL) passou a ser estatisticamente significativa após a eliminação das variáveis *Filhos* e *Nota de Ingresso*. Assim, pode-se concluir que, para prever os abates, a

especialidade do militar deve ser considerada, sendo que as identificadas pelo modelo RL 1.0 são aquelas com maior propensão ao abate, levando em conta as demais condições do modelo. Isto não implica que outras especialidades não possam solicitar abate aos QP, mas indica que a FA deve prestar especial atenção a estas.

No que diz respeito à variável *Idade*, verifica-se que o aumento da idade influencia a probabilidade de saída do militar, sendo esse impacto positivo, o que significa que à medida que a idade aumenta a probabilidade de saída é maior.

Por outro lado, a variável *Tempo de Serviço Efetivo* apresenta um comportamento inverso, ou seja, à medida que o tempo de serviço efetivo aumenta, diminuiu a probabilidade de pedido de abate aos QP.

O *Tempo de Permanência no Posto* também tem um impacto positivo na intenção de pedir abate aos QP, nomeadamente quanto maior o tempo de permanência num determinado posto, maior será a probabilidade de saída voluntária de um militar dos QP. Neste sentido, é importante encontrar mecanismos que garantam uma transição mais rápida entre os postos, ou no caso dos postos em que o tempo de permanência é mais elevado, por força das disposições estatutárias previstas no EMFAR, deve-se encontrar mecanismos que atenuem esta intenção de saída.

Em relação à *Nota FAV*, observa-se que o aumento no seu valor tem um impacto positivo na probabilidade de ficar nos QP, ou seja, pode ser um bom indicador para a retenção dos militares dos QP. Por outro lado, este resultado pode sugerir que os militares que solicitam abate podem sentir que seu trabalho não é devidamente reconhecido através dos resultados da sua FAV, o que poderá levar à consideração de um caminho alternativo, conforme defendido por Lee & Mitchell (1994).

O número de louvores também apresenta um resultado significativo no modelo, sugerindo que um aumento nesse valor tem um impacto positivo na decisão de saída. Por outras palavras, se um militar receber um louvor, a probabilidade de sair aumenta. À primeira vista, isso pode parecer contraditório. No entanto, essa aparente contradição pode ser compreendida sob a ótica da retenção de talentos. Os militares que recebem mais louvores são frequentemente aqueles que se destacam por "atos ou comportamentos que revelem notáveis valores, competência profissional, entrega ao cumprimento dos deveres ou civismo", conforme descrito na Lei Orgânica n.º 2/2009 - Regulamento de Disciplina Militar (p. 4617). Esses militares são, portanto, indivíduos cujo talento e competência são amplamente reconhecidos pela organização. Assim, o facto de militares com mais louvores terem uma maior probabilidade de pedir abate pode refletir uma questão de atração de talentos por outras oportunidades fora da instituição. Como profissionais altamente competentes e reconhecidos, esses militares podem estar mais aptos a procurar, ou a

receber, ofertas externas que valorizem ainda mais as suas competências. Portanto, é crucial que a organização reconheça que o talento mais valorizado pode ser também o mais difícil de reter, exigindo estratégias específicas para manter esses profissionais na FA.

Por fim, é importante destacar que o número de formações tem um impacto negativo na probabilidade de saída dos militares dos QP. Isto significa que, à medida que o militar completa mais formações, a probabilidade da sua saída diminui, indicando que a formação é um dos elementos mais valorizados pelos militares.

Desta forma, o modelo RL 1.0 permite atingir um dos objetivos específicos definidos, que é identificar os indicadores (variáveis) nos sistemas de informação da área de RH da FA que podem influenciar as saídas. Foram identificadas 10 variáveis que afetam a decisão de um militar em solicitar abate aos QP. Esses indicadores são: Idade, Ensino Superior, Sexo Masculino, Nota FAV, Número de Louvores, Número de Formações, Especialidade (ENGAER, ENGEL, PILAV, MED, OPCART, MMA, MELIAV, MARME), Posto (TCOR), Tempo no Posto e Tempo de Serviço Efetivo. Para responder ao segundo objetivo específico, que era a definição de um índice de risco de saída voluntária dos militares dos QP, também se poderá utilizar, para esse efeito, os resultados do modelo RL 1.0, através do resultado obtido da sua probabilidade de saída.

Ao utilizar a equação da RL, que no final resulta num valor entre 0 e 1 que representa a probabilidade de cada militar dos QP pedir abate aos QP, ou seja, valores acima de 0,5 representam uma saída voluntária dos QP. Ao multiplicar esse valor por 100, obtemos um índice associado à probabilidade de saída de cada militar.

A equação da RL é representada pela equação:

$$\log\left(\frac{\rho}{1-\rho}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad [1]$$

Onde $\log\left(\frac{\rho}{1-\rho}\right)$ é o logaritmo das probabilidades de sucesso, ou seja, a probabilidade de $Y = 1$ contra $Y=0$, β_0 é a constante, $\beta_1, \beta_2, \beta_n$ são os coeficientes das variáveis preditoras e X_1, X_2, X_n são as variáveis preditoras.

No caso concreto do modelo RL 1.0 obtém-se a seguinte equação:

$$\log\left(\frac{\rho}{1-\rho}\right) = -10,86 + 0,47 * \text{Idade} + 1,29 * \text{Ensino Superior} + 35,58 * 2SAR + 2,88 * ALF + (-1,95) * CAP + (-20,33) * COR + (-1,23) * MAJ + 0,08 * SAJ + \quad [2]$$

$$\begin{aligned}
& (-0,55) * SCH + (-15,68) * SMOR + (-4,02) * \mathbf{TCOR} + 1,30 * TEN + 0,16 * \\
& \mathbf{tmp_{posto}} + 0,19 * \mathbf{Louvores} + -0,1 * \mathbf{Formações} + 0,92 * \mathbf{Masculino} + (-0,41) * \\
& \mathbf{Tmp_{srvefe}} + (-14,90) * ADMAER + 0,19 * CMI + 3,07 * ENGAED + 5,56 * \mathbf{ENGAER} + \\
& 3,33 * \mathbf{ENGEL} + (-15,96) * JUR + 1,99 * \mathbf{MARME} + 3,46 * \mathbf{MED} + 0,36 * \\
& MELECA + (-16,13) * MELECT + 2,38 * \mathbf{MELIAV} + 2,86 * \mathbf{MMA} + (-0,27) * \\
& MMT + (-1,94) * MUS + 1,33 * NAV + 2,84 * \mathbf{OPCART} + (-1,04) * OPCOM + 0,44 * \\
& OPINF + (-16,64) * OPMET + 0,30 * OPRDET + 0,25 * OPSAS + 0,73 * PA + \\
& 4,55 * \mathbf{PILAV} + 2,83 * PSI + (-1,62) * SAS + (-16,88) * TABST + 1,80 * TINF + \\
& (-16,70) * TMAEQ + (-16,63) * TMI + 2,18 * TMMA + 1 * TMMEL + 1,02 * \\
& TMMT + 0,10 * TOCART + (-16,71) * TOCC + (-16,38) * TODCI + (-16,50) * \\
& TOMET + (-0,75) * TPAA + (-16,63) * TS + (-1,08) * \mathbf{Nota_{fav}}
\end{aligned}$$

Desta forma, o resultado obtido para cada um dos militares, pela aplicação desta equação e multiplicado por 100, resulta num índice para cada um deles, com a sua probabilidade em pedir abate aos QP, atingindo-se assim o outro objetivo específico.

Perante isto, considera-se que o objetivo geral desta investigação foi atingido, uma vez que se conseguiu estudar a problemática dos abates dos militares dos QP da FA, utilizando os dados existentes nos sistemas de informação de RH da instituição.

É possível então responder à pergunta de partida, na medida em que se consegue prever as saídas dos militares dos QP da FA com base nos dados disponíveis sobre eles. No entanto, esta previsão precisa de ser afinada ao longo dos anos, com a inclusão de outras variáveis, por exemplo, uma vez que se verificou que só com as variáveis utilizadas, o modelo não conseguiu identificar, com a sensibilidade desejada, os militares que pediram abate aos QP, o que veio comprovar a complexidade do problema da rotação e retenção de talentos, conforme apontado por autores como Chiavenato (2014), Halim et al. (2020), Hom et al. (2017) e Lee & Mitchell (1994).

Porém, considera-se que a investigação realizada apresenta um contributo válido para o estudo da problemática por parte da FA, com a identificação de alguns fatores que podem contribuir para os abates aos QP, podendo ser criados mecanismos para tentar mitigar essa situação.

Conclusão

Os Recursos Humanos são, por norma, o bem mais valioso de uma Organização, uma vez que lhe permite ter inteligência na gestão dos negócios e a racionalidade nas decisões, destacando-se assim das demais (Chiavenato, 2014; Halim et al., 2020).

Para Chiavenato (2014), a Era da Informação modificou a forma com o capital humano era visto pelas organizações, no sentido em que se percebeu que este bem organizacional passou a ser a nova moeda dos negócios, através da gestão dos seus talentos.

De facto, a retenção de colaboradores tornou-se num dos grandes desafios organizacionais do século XXI. As organizações perceberam que os custos associados à saída voluntária de um colaborador são extremamente elevados. Isso ocorre porque, ao sair, o colaborador leva consigo todo o conhecimento e experiência adquiridos, o que representa uma perda significativa para a organização. Além disso, a substituição deste colaborador geralmente implica custos inesperados (Nunes et al., 2021).

Para fazer face a esta situação, as organizações têm procurado entender os motivos pelos quais os seus colaboradores optam por sair voluntariamente, através de estudos focados na retenção de talento e consequente análise do *turnover*. O objetivo dessas investigações é, assim, procurar identificar as causas que levam à saída dos seus colaboradores, permitindo que sejam desenvolvidos mecanismos e estratégias para mitigar essas saídas (Halim et al., 2020).

No caso das FFAA, o problema da retenção dos militares dos QP tem sido uma preocupação para os últimos governos, que têm observado uma redução gradual desses efetivos, impulsionada pelo aumento dos pedidos de abate aos QP. A saída contínua de militares qualificados representa um desafio significativo, não apenas para a manutenção da capacidade operacional das Forças Armadas, mas também para a preservação do conhecimento e da experiência acumulados por esses profissionais ao longo dos anos. Como resultado, o problema da retenção tornou-se uma prioridade estratégica, exigindo medidas eficazes para reter talentos e evitar uma maior erosão do capital humano nas FFAA.

Esta investigação centrou-se na problemática das saídas voluntárias dos militares dos QP da FA, um desafio significativo para a GRH em contexto militar, conforme reconheceu o Chefe do Estado-Maior da Força Aérea, General João Cartaxo Alves (2024).

Assim, o objetivo principal foi desenvolver um modelo preditivo capaz de identificar quais os militares com maior probabilidade de pedirem abate aos QP, utilizando para o efeito os dados disponíveis nos sistemas de informação de RH da instituição.

Através da aplicação de diversos algoritmos de *machine learning*, orientados para a metodologia da classificação, foi possível construir um modelo de RL que apresentou bons resultados em

termos de exatidão e precisão, embora com algumas limitações no que diz respeito à sensibilidade.

A análise do fenómeno do abate aos QP foi considerada fundamental para a FA, que, nos últimos anos, tem enfrentado uma crescente dificuldade em manter os seus efetivos, em parte devido ao aumento dos pedidos de abate aos QP. A problemática da rotatividade de pessoal não é exclusiva das FFAA, mas assume características próprias devido à natureza da carreira militar e às exigências específicas da vida castrense.

A substituição de um militar com anos de experiência e formação não é apenas dispendiosa, mas também demorada, dado que envolve o recrutamento, o treino e a adaptação dos novos militares ao ambiente militar (Alves, 2024).

O presente estudo teve como ponto de partida a análise de uma base de dados composta por 2176 militares, dos quais 1993 não pediram abate e 183 solicitaram a saída entre 2019 e 2023. As variáveis utilizadas foram divididas em dois grupos: sociodemográficas e estruturais, permitindo assim uma visão abrangente dos fatores que poderiam influenciar a decisão de saída dos militares. As variáveis sociodemográficas incluíram o sexo, a idade, o estado civil, o número de filhos, a tipologia da área de residência e habilitações literárias. As variáveis estruturais focaram-se em aspetos relacionados com a carreira militar, como o posto, a especialidade, a categoria, o número de louvores recebidos, o número de condecorações recebidas, a nota da FAV, o número de formações frequentadas, o tempo de serviço efetivo, o tempo de QP, o tempo no posto, o tempo de colocação na Unidade, o número de colocações, o número de funções desempenhadas, a distância entre a unidade de colocação e a residência, a nota de ingresso nos QP.

A operacionalização da investigação seguiu uma metodologia baseada na aplicação de cinco métodos de classificação supervisionada: Regressão Logística, *Naïve Bayes*, Análise Discriminante Linear, Análise Discriminante Quadrática e Árvores de Decisão. Estes cinco métodos foram analisados e avaliados através da utilização da matriz de confusão, concretamente pela análise de cinco métricas: exatidão, precisão, sensibilidade, especificidade e *F1-Score*. No final optou-se pela escolha do modelo de RL que apresentou valores mais equilibrados nos cinco fatores avaliados e na curva ROC.

Importa realçar que todos os métodos analisados apresentaram valores válidos, mas a escolha acabou por recair na RL por se considerar mais adequado para o contexto da investigação e dos objetivos propostos.

O modelo de RL foi então aperfeiçoado através da técnica de *Stepwise*, recorrendo ao critério AIC e aos valores de *p-value* inferiores a 0,05, para a seleção das variáveis. Desta forma, o

modelo inicial, que contava com 21 variáveis, foi reduzido para um modelo denominado RL 1.0, contendo apenas 10 indicadores. Este aperfeiçoamento resultou em melhorias nos cinco valores analisados pela matriz de confusão. No entanto, a sensibilidade do modelo — capacidade de identificar corretamente os militares que solicitaram abate — permaneceu relativamente baixa, indicando a necessidade de melhoramentos futuros.

Em termos de resultados, o estudo permitiu identificar um conjunto de variáveis que se destacaram pela sua influência na decisão de saída voluntária dos militares dos QP. A idade revelou-se uma das variáveis mais significativas, com uma relação positiva com a probabilidade de abate, ou seja, à medida que os militares envelhecem, a probabilidade de pedirem abate aumenta. Por outro lado, o tempo de serviço efetivo demonstrou ter um impacto inverso, ou seja, quanto mais tempo o militar permanecer na efetividade de serviço, menor é a probabilidade de pedir abate. Este resultado é consistente com a carreira militar, onde o tempo de serviço é frequentemente associado à progressão na hierarquia e a um maior compromisso com a organização (Machado & Fachada, 2018; Mehl et al., 2022).

A questão do tempo de permanência no posto também foi uma variável considerada significativa na decisão de abate, tendo-se verificado que quanto maior o tempo de permanência no posto, maior a probabilidade de saída.

Outra conclusão relevante foi o impacto que o número de louvores recebidos tem na decisão de saída. Surpreendentemente, os militares que receberam mais louvores apresentaram uma maior probabilidade de pedir abate. Este resultado contradiz a expectativa de que o reconhecimento institucional, através de louvores, funcione como um fator de retenção. No entanto, pode ser interpretado à luz da gestão de talentos, uma vez que os militares com mais louvores tendem a ser aqueles com maior visibilidade e, possivelmente, com mais oportunidades fora da organização, o que pode levá-los a considerar uma saída voluntária.

Por outro lado, o número de formações frequentadas mostrou ter um impacto negativo na probabilidade de abate, sugerindo que os militares valorizam as oportunidades de formação e desenvolvimento oferecidas pela FA, e que estas podem funcionar como um fator de retenção eficaz.

A nota da última FAV também apresentou um efeito significativo na decisão de saída dos militares. Aqueles com melhores notas na FAV mostraram uma menor probabilidade de pedir abate, o que reforça a importância de um sistema de avaliação justo e meritocrático. Os militares que sentem que o seu desempenho é devidamente reconhecido tendem a apresentar maior satisfação com a organização e, conseqüentemente, uma menor predisposição para sair.

Por fim, o estudo permitiu desenvolver um índice de probabilidade de abate aos QP, baseado nos resultados do modelo RL1.0. Este índice pode ser utilizado pela FA para monitorizar, em tempo real, os militares com maior risco de pedirem abate, permitindo assim uma gestão proativa dos recursos humanos. A capacidade de prever as saídas voluntárias pode ser integrada nas políticas de gestão de recursos humanos da FA, permitindo o desenvolvimento de estratégias de retenção mais direcionadas e eficazes.

Apesar dos contributos significativos deste estudo, é importante reconhecer as suas limitações. Em primeiro lugar, referir que a base de dados utilizada incluía apenas um conjunto restrito de variáveis, o que pode ter limitado a capacidade do modelo de captar toda a complexidade envolvida na decisão de saída voluntária dos militares. Por exemplo, fatores externos como o ambiente familiar, as condições socioeconómicas ou o equilíbrio entre vida pessoal e profissional, que podem ter um impacto considerável na decisão de pedido de abate, não foram considerados. No entanto, convém destacar que o objetivo deste trabalho era prever os abates com base nos dados disponíveis nos sistemas de informação dos serviços de recursos humanos da FA, inviabilizando, assim, a inclusão de variáveis externas. Em segundo lugar, o valor relativamente baixo da sensibilidade do modelo sugere que alguns casos de abate não foram corretamente identificados, o que aponta para a possibilidade de melhoria, seja pela inclusão de mais variáveis, seja pela divisão dos militares por categoria (Oficiais/Sargentos).

Relativamente aos contributos práticos, este estudo oferece uma ferramenta válida para a GRH da FA, ao permitir identificar precocemente os militares com maior probabilidade de pedirem abate. Com base nos resultados do modelo RL1.0, a FA pode desenvolver estratégias de retenção mais eficazes, focadas nas variáveis que mais influenciam a decisão de saída, como a oferta de mais formações, o reconhecimento adequado do desempenho e a melhoria das condições de trabalho. A retenção de militares experientes e bem formados é crucial para a manutenção da capacidade operacional da FA, e este estudo fornece uma base sólida para a implementação de políticas de retenção mais eficazes.

Ainda dentro do modelo de RL e para melhorar a sua capacidade preditiva, uma abordagem eficaz seria incluir interações entre variáveis independentes. A introdução de termos de interação permite que o modelo capture efeitos combinados entre variáveis que, quando analisadas isoladamente, podem não refletir totalmente a sua influência sobre a variável dependente. Por exemplo, a interação entre a idade e o tempo de serviço efetivo pode revelar que o impacto da idade na probabilidade de saída dos militares se altera dependendo do tempo de serviço acumulado, captando nuances que variáveis individuais não conseguiriam identificar isoladamente. Este refinamento pode melhorar a sensibilidade e especificidade do modelo,

aumentando a sua capacidade de identificar corretamente os militares com maior probabilidade de solicitar abate, além de proporcionar uma compreensão mais profunda das relações entre fatores que influenciam a retenção e a rotatividade. Ao incluir interações, o modelo torna-se mais robusto, capaz de refletir a complexidade da tomada de decisão dos militares e contribuindo para previsões mais precisas.

Outra linha de investigação possível seria a aplicação de algoritmos mais avançados, como redes neurais ou técnicas de *deep learning*, que podem ter um melhor desempenho na identificação de padrões complexos nos dados. Finalmente, poderia ser avaliado o impacto das estratégias de retenção baseadas nos resultados deste modelo, quando implementadas, de modo a verificar se estas são eficazes na redução das saídas voluntárias dos militares dos QP.

Este estudo contribui para a literatura sobre rotatividade e retenção de talentos no contexto militar, ao identificar os principais fatores que influenciam a decisão de pedido de abate dos militares dos QP e ao desenvolver um modelo preditivo capaz de antecipar estes abates. Os resultados obtidos nesta investigação têm implicações práticas na GRH da FA, e abrem caminho para futuras investigações que possam aprofundar o conhecimento sobre este fenómeno e contribuir para a criação de estratégias de retenção mais eficazes.

Referências bibliográficas

- Agência para a Modernização Administrativa. (2020). *Tipologia das áreas urbanas (2014)*. <https://dados.gov.pt/pt/datasets/tipologia-das-areas-urbanas-2014/>
- Alves, J. (2024, Abril 12). *General Cartaxo Alves: “Tenho militares que chegam a pagar 150 mil euros para se desvincularem”*. Diário de Notícias. <https://www.dn.pt/2235481395/general-cartaxo-alves-tenho-militares-que-chegam-a-pagar-150-mil-euros-para-se-desvincularem/>
- Bosdotter, S., Wellborg, T., & Sandström, J. (2022). Why Fighter Pilots are Leaving the Swedish Armed Forces – and how to Retain them. *Scandinavian Journal of Military Studies*, 5(1), 364–378. <https://doi.org/10.31374/sjms.135>
- Branco, P. (2022). *Problemática do Turnover no Exército e a Importância da Retenção de Talentos Escola Superior de Gestão de Tomar Problemática do Turnover no Exército e a Importância da Retenção de Talentos*.
- Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. P. (2020). SMOTE: Synthetic Minority Over-sampling Technique Nitesh. *Journal of Artificial Intelligence Research*, 4(1), 321–357.
- Chiavenato, I. (2014). *Gestão de Pessoas - O novo papel dos recursos humanos nas organizações* (4ª Edição). Editora Manole.
- Congard, A., Delicourt, A., Dauvier, B., & Gros, F. (2022). Personality trait and subjective work experience fit (P-SWE fit) as an explanation for voluntary turnover. *Military Psychology*, 34(1), 68–82. <https://doi.org/10.1080/08995605.2021.1970459>
- Decreto-Lei n.º 90/2015 Estatuto dos Militares das Forças Armadas, (2015). <https://diariodarepublica.pt/dr/legislacao-consolidada/decreto-lei/2015-114793591>
- Força Aérea. (2022). *Planeamento Estratégico Força Aérea - 2022-2025*.
- Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (1ª edição). Cambridge University Press.
- Gultekin, Z., Abdan, E., & Kilic, S. Y. (2012). *A Military Application of Human Capital Management: Military Pilot Turnover*. «Henri Coanda» - Romania Air force Academy.

- Halim, Z., Maria, Waqas, M., Edwin, C. A., & Shah, A. (2020). Identifying factors for employee retention using computational techniques: an approach to assist the decision-making process. *SN Applied Sciences*, 2(9), 1–20. <https://doi.org/10.1007/s42452-020-03415-5>
- Hayat Suhendar, M. T., & Widyani, Y. (2023). Machine Learning Application Development Guidelines Using CRISP-DM and Scrum Concept. *Proceedings of 2023 IEEE International Conference on Data and Software Engineering, ICoDSE 2023*, 168–173. <https://doi.org/10.1109/ICoDSE59534.2023.10291438>
- Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology*, 102(3), 530–545. <https://doi.org/10.1037/apl0000103>
- Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). Applied Logistic Regression. Em *Biometrics* (3ª Edição, Vol. 47, Número 4). John Wiley & Sons, Inc. <https://doi.org/10.2307/2532419>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An Introduction to Statistical Learning, with applications in R* (2ª edição). Springer.
- Lantz, B. (2013). Machine Learning with R. Em *Machine Learning* (1ª edição). Packt Publishing Ltd. <https://doi.org/10.1002/9781119183464.ch12>
- Lee, T. W., & Mitchell, T. R. (1994). An Alternative Approach: The Unfolding Model of Voluntary Employee Turnover. *The Academy of Management Review*, 19(1), 51. <https://doi.org/10.2307/258835>
- Lusa. (2023a). *Força Aérea Portuguesa precisa de mais 1400 militares, diz chefe do Estado-Maior*. CNN Portugal. <https://cnnportugal.iol.pt/forca-aerea-portuguesa/chefe-do-estado-maior/forca-aerea-portuguesa-precisa-de-mais-1-400-militares/20230701/64a04b2ad34e3ae5b8c340da>
- Lusa. (2023b). *Forças Armadas com 23.425 militares e efetivos: “Drasticamente abaixo do autorizado*. Público. <https://www.publico.pt/2023/10/31/politica/noticia/forcas-armadas-23425-militares-efetivos-drasticamente-abaixo-autorizado-2068674>
- Lusa. (2024). *Nuno Melo: reuniões com militares para rever carreira já começaram e*

- haverá medidas «muito brevemente».* Expresso.
<https://expresso.pt/politica/defesa/2024-07-16-nuno-melo-reunioes-com-militares-para-rever-carreira-ja-comecaram-e-havera-medidas-muito-brevemente-c59f4f8b>
- Machado, J., & Fachada, C. (2018). A POLÍTICA DE INDEMNIZAÇÕES NA FORÇA AÉREA PORTUGUESA : UMA MEDIDA DE RETENÇÃO DOS THE PORTUGUESE AIR FORCE ' S INDEMNIFICATION POLICY : A. *Revista de Ciências Militares*, VI (2), 379–402. <https://www.ium.pt/cisdi/index.php/pt/publicacoes/revista-de-ciencias-militares>.
- Mafini, C., & Dubihlela, J. (2013). Determinants of military turnover of technical air-force specialists: An empirical case analysis. *Mediterranean Journal of Social Sciences*, 4(3), 523–534. <https://doi.org/10.5901/mjss.2013.v4n3p523>
- Maroco, J. (2021). *Análise Estatística com o SPSS Statistics* (8ª Edição). Edições Sílabos.
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>
- Martinez, I., Viles, E., & G. Olaizola, I. (2021). Data Science Methodologies: Current Challenges and Future Approaches. *Big Data Research*, 24, 100183. <https://doi.org/10.1016/j.bdr.2020.100183>
- Martins, A. M. de B. S. Q. (2014). *Turnover E Retenção Dos Militares Do Regime De Contrato Da Força Aérea* [Lisboa School of Economics & Management]. <https://www.repository.utl.pt/handle/10400.5/7714>
- Marvin, G., Jackson, M., & Alam, M. G. R. (2021). A Machine Learning Approach for Employee Retention Prediction. *TENSYMP 2021 - 2021 IEEE Region 10 Symposium, July 2023*. <https://doi.org/10.1109/TENSYMP52854.2021.9550921>
- Mazlan, M. R. M., & Jambulingam, M. (2023). Challenges of Talent Retention: A Review of Literature. *Journal of Business and Management Review*, 4(2), 078–091. <https://doi.org/10.47153/jbmr42.6302023>

- McMahon, C. J., & Bernard, C. J. (2018). Storm Clouds on the Horizon. *Naval War College Review*, 72, 84–99. <https://doi.org/10.2307/j.ctv6wgc2v.14>
- Mehl, D., Faruki, A., & Ehinger, J. (2022). Using advanced analytics to turn attrition into retention. *Supply Chain Management Review*, 26(1), 42–45. <https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=bth&AN=154694962&site=ehost-live&custid=s1020214>
- Meirinhos, M., & Osório, A. (2010). Educação O estudo de caso como estratégia de investigação em educação The case study as research strategy in education. *Revista de Educação*, 2(2), 49–65.
- Meirinhos, R., & Barreto, A. (2018). A Comunicação Estratégica como Fator de Retenção de Recursos Humanos. *Media & Jornalismo*, 18(33), 75–90. https://doi.org/10.14195/2183-5462_33_5
- Ministério Defesa Nacional. (2023). *Plano De Ação Para a Profissionalização Do Serviço Militar*.
- Mitchell, T. (1997). Machine Learning. Em *Natural Computing Series*. McGraw-Hill Science/Engineering/Math. https://doi.org/10.1007/978-3-031-17922-8_9
- Mohamed, A. E. (2017). Comparative Study of Four Supervised Machine Learning Techniques for Classification. *International Journal of Applied Science and Technology*, 7(2), 5–18.
- Morrel, K., Loan-Clarke, J., & Wilkinson, A. (2001). Lee And Mitchell's Unfolding Model of Employee Turnover – A Theoretical Assessment. *Business School Research Series*, 1–40.
- Munish, & Agarwal, R. (2017). Employee Engagement & Retention: A Review of Literature. *International Journal of BRIC Business Research*, 6(1), 01–19. <https://doi.org/10.14810/ijbbr.2017.6101>
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons.B*, 4(December), 51–62. <https://doi.org/10.20544/horizons.b.04.1.17.p05>
- Neves, P. (2023). *A retenção dos militares na Marinha Portuguesa* [Escola Naval].

- <https://comum.rcaap.pt/bitstream/10400.26/49043/1/507> - ASPOF AN Brandão Neves.pdf
- Nunes, A. L. D. P. F., Martins, G. G., & Mendonça, J. D. M. (2021). A Retenção de Talentos e o Novo Normal de Recursos Humanos / Talent Retention and the New Normal of Human Resources. *ID on line. Revista de psicologia*, 15(58), 391–409. <https://doi.org/10.14295/online.v15i58.3335>
- Resolução do Conselho de Ministros n.º 19/2013: Conceito Estratégico Defesa Nacional, (2013). <https://dre.pt/dre/detalhe/resolucao-conselho-ministros/19-2013-259967>
- Rijo, F., Marreiros, J., Mairos, J., & Paquete, O. (2018). a Retenção Dos Militares Nas Forças Armadas. *Revista de Ciências Militares*, 1, 333–356. <https://www.ium.pt/cisdi/index.php/pt/publicacoes/revista-de-ciencias-militares>.
- Saltz, J. S. (2021). CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps. *Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021*, 2337–2344. <https://doi.org/10.1109/BigData52589.2021.9671634>
- Santos, L. A. B. (2021). *As Forças Armadas Portuguesas no mercado de trabalho: análise crítica do atual modelo de prestação de serviços militar em regime de contrato* [Universidade Lusíada]. <http://hdl.handle.net/11067/6054>
- Santos, L., & Lima, J. (2019). Orientações metodológicas para a elaboração de trabalhos de investigação. *Cadernos do IUM*, 8.
- Sims, C. S., Trail, T. E., & Coe, J. (2024). Predicting soldier retention from army spouse characteristics and attitudes: Soldiering on with spouse support. *Military Psychology*, 00(00), 1–12. <https://doi.org/10.1080/08995605.2024.2319014>
- Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. *Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016*, 1310–1315.
- Souza, P. L., Casañeda-Ayarza, J., & Spilleir, D. (2022). Retenção de Talentos e Turnover: Diagnóstico e Discussão de Fatores a Partir de um Estudo de Caso. *Revista de Administração IMED*, 128–151. <https://doi.org/https://doi.org/10.18256/2237-7956.2022.v12i1.4448>

- Taleb, N. (2011). *A Lógica do Cisne Negro: o impacto do altamente improvável* (5ª edição). Dom Quixote.
- Vrigazova, B. (2021). The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems. *Business Systems Research*, 12(1), 228–242. <https://doi.org/10.2478/bsrj-2021-0015>
- Yedida, R., Vahhi, R., Abhilash, Reddy, R., J, R., & Kilkarni, D. (2024). Employee Attrition Prediction. *Indian Journal of Data Mining*, 4(1), 26–30. <https://doi.org/10.54105/ijdm.a1636.04010524>
- Yin, R. K. (2001). *ESTUDO DE CASO PLANEAMENTO E MÉTODOS* (2ª Edição). Bookman.

Apêndices

Apêndice 1 – Matriz de confusão 5 métodos

		RL	
		PREVISTO	
		Não	Sim
REAL	Não	588	10
	Sim	37	18

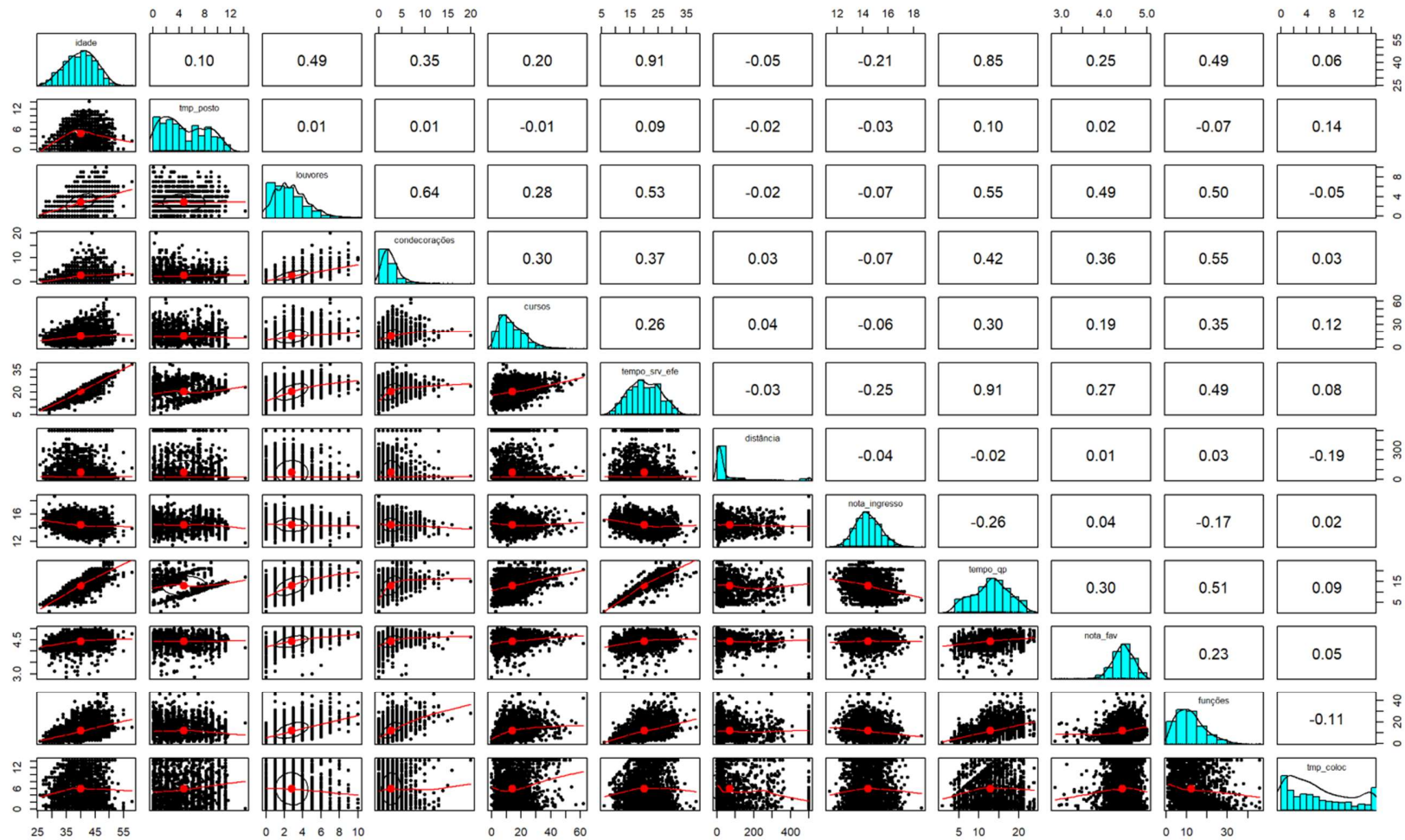
		AD	
		PREVISTO	
		Não	Sim
	Não	591	7
	Sim	47	8

		NB	
		PREVISTO	
		Não	Sim
REAL	Não	514	84
	Sim	31	24

		LDA	
		PREVISTO	
		Não	Sim
	Não	586	12
	Sim	39	16

		QLA	
		PREVISTO	
		Não	Sim
REAL	Não	578	20
	Sim	47	8

Apêndice 2 – Gráficos das estatísticas descritivas



Apêndice 3 – Resultado das variáveis do modelo inicial (21 variáveis)

Variável	Fatorial	Coefficiente	Erro est.	Estatística	P-value
(Constante)		-7,850	10972,060	-0,001	0,999
Filhos		-0,212	0,146	-1,446	0,148
Idade		0,462	0,065	7,102	0,000 ***
Tipo Área	APR	0,168	0,523	0,322	0,747
Residência	APU	0,098	0,376	0,261	0,794
Habilitações	Ensino Superior	1,349	0,416	3,245	0,001 ***
Posto	2SAR	34,864	4997,803	0,007	0,994
	ALF	3,965	15731,380	0,000	1,000
	CAP	-0,766	10972,050	0,000	1,000
	COR	-18,757	11325,710	-0,002	0,999
	MAJ	0,111	10972,050	0,000	1,000
	SAJ	0,409	1,556	0,263	0,793
	SCH	0,017	2,648	0,006	0,995
	SMOR	-15,203	10754,010	-0,001	0,999
	TCOR	-2,577	10972,050	0,000	1,000
	TEN	2,540	10972,050	0,000	1,000
Tempo Posto		0,181	0,154	1,175	0,240
Nº Louvores		0,210	0,109	1,925	0,054 .
Nº Condecorações		0,033	0,087	0,383	0,702
Nº Formações		-0,091	0,023	-3,948	0,000 ***
Sexo	Masculino	0,936	0,430	2,177	0,029 *
Categoria	Sargentos	0,943	10972,050	0,000	1,000
Tmp Serviço Efetivo		-0,408	0,080	-5,090	0,000 ***
Distância		0,000	0,001	0,336	0,737
Nota Ingresso		-0,214	0,137	-1,565	0,118
Estado Civil	Divorciado	-0,510	0,556	-0,917	0,359
	Solteiro	-0,050	0,277	-0,181	0,856
	Unido de facto	-0,348	0,369	-0,943	0,346
	Viúvo	-15,827	10754,010	-0,001	0,999
Especialidade	ADMAER	-15,248	1740,649	-0,009	0,993
	CMI	-0,115	1,331	-0,086	0,931
	ENGAED	2,414	1,649	1,464	0,143
	ENGAER	5,039	1,562	3,227	0,001 **
	ENGEL	2,723	1,589	1,714	0,087 .
	JUR	-15,695	4917,808	-0,003	0,997
	MARME	1,892	0,971	1,948	0,051 .
	MED	3,221	1,562	2,062	0,039 *
	MELECA	0,138	0,986	0,140	0,889
	MELECT	-16,254	1621,692	-0,010	0,992
	MELIAV	2,190	0,978	2,238	0,025 **
	MMA	2,683	0,891	3,012	0,003 **
	MMT	-0,554	1,322	-0,419	0,675
	MUS	-2,192	1,389	-1,578	0,115
	NAV	1,435	1,813	0,792	0,429
	OPCART	2,478	1,035	2,393	0,017 *
	OPCOM	-1,200	1,303	-0,921	0,357
	OPINF	0,249	1,338	0,186	0,853
	OPMET	-16,700	2176,505	-0,008	0,994
	OPRDET	0,178	1,327	0,134	0,894
	OPSAS	0,256	1,340	0,191	0,848
	PA	0,723	1,039	0,696	0,487
	PILAV	4,064	1,492	2,723	0,006 **
	PSI	2,272	1,808	1,256	0,209
	SAS	-1,622	1,108	-1,465	0,143
	TABST	-16,796	2424,731	-0,007	0,994

	TINF	1,621	1,505	1,077	0,282
	TMAEQ	-16,950	3382,520	-0,005	0,996
	TMI	-16,870	5065,261	-0,003	0,997
	TMMA	2,149	1,525	1,409	0,159
	TMMEL	0,876	1,733	0,505	0,613
	TMMT	0,824	1,754	0,470	0,639
	TOCART	-0,095	1,726	-0,055	0,956
	TOCC	-16,721	3380,155	-0,005	0,996
	TODCI	-16,508	2002,203	-0,008	0,993
	TOMET	-16,465	3579,965	-0,005	0,996
	TPAA	-0,802	1,717	-0,467	0,641
	TS	-16,790	3290,542	-0,005	0,996
Tempo QP		-0,012	0,184	-0,068	0,946
Nota Avaliação		-1,194	0,490	-2,434	0,015 *
Funções		-0,030	0,026	-1,144	0,253
Tmp Colocação		0,014	0,036	0,388	0,698
Colocações		0,014	0,105	0,133	0,894

p-value: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1

Apêndice 4 – Resultado das variáveis do modelo RL 1.0

Variável	Fatorial	Coefficiente	Erro est.	Estatística	P-value
(Constante)		-10,860	2,955	-3,676	0,000 ***
Idade		0,468	0,064	7,346	0,000 ***
Habilitações	Ensino Superior	1,287	0,408	3,152	0,002 **
Posto	2SAR	35,580	5128,000	0,007	0,994
	ALF	2,880	11280,000	0,000	1,000
	CAP	-1,948	1,117	-1,744	0,081
	COR	-20,330	2807,000	-0,007	0,994
	MAJ	-1,233	1,331	-0,926	0,354
	SAJ	0,083	0,787	0,106	0,916
	SCH	-0,550	1,646	-0,334	0,738
	SMOR	-15,680	10750,000	-0,001	0,999
	TCOR	-4,020	1,743	-2,306	0,021 *
	TEN	1,297	2,344	0,553	0,580
Tempo Posto		0,160	0,080	1,993	0,046 *
Nº Louvores		0,193	0,093	2,066	0,039 *
Nº Formações		-0,101	0,023	-4,397	0,000 ***
Sexo	Masculino	0,922	0,421	2,189	0,029 *
Tmp Serviço Efetivo		-0,409	0,072	-5,704	0,000 ***
Especialidade	ADMAER	-14,900	1748,000	-0,009	0,993
	CMI	0,193	1,309	0,147	0,883
	ENGAED	3,067	1,609	1,906	0,057
	ENGAER	5,556	1,538	3,614	0,000 ***
	ENGEL	3,329	1,542	2,158	0,031 *
	JUR	-15,960	4787,000	-0,003	0,997
	MARME	1,997	0,954	2,094	0,036 *
	MED	3,461	1,540	2,248	0,025 *
	MELECA	0,355	0,966	0,367	0,713
	MELECT	-16,130	1621,000	-0,010	0,992
	MELIAV	2,377	0,958	2,482	0,013 *
	MMA	2,864	0,870	3,290	0,001 **
	MMT	-0,269	1,302	-0,207	0,836
	MUS	-1,939	1,336	-1,451	0,147
	NAV	1,325	1,768	0,749	0,454
	OPCART	2,835	0,991	2,860	0,004 **
	OPCOM	-1,038	1,289	-0,805	0,421
	OPINF	0,439	1,320	0,333	0,739
	OPMET	-16,640	2168,000	-0,008	0,994
	OPRDET	0,299	1,312	0,228	0,820
	OPSAS	0,254	1,327	0,191	0,849
	PA	0,729	1,024	0,711	0,477
	PILAV	4,555	1,468	3,104	0,002 **
	PSI	2,083	1,774	1,174	0,240
	SAS	-1,615	1,094	-1,476	0,140
	TABST	-16,870	2369,000	-0,007	0,994
	TINF	1,802	1,491	1,209	0,227
	TMAEQ	-16,700	3391,000	-0,005	0,996
	TMI	-16,630	5058,000	-0,003	0,997
	TMMA	2,175	1,511	1,439	0,150
	TMMEL	1,001	1,710	0,586	0,558
	TMMT	1,018	1,726	0,590	0,556
	TOCART	0,104	1,703	0,061	0,951
	TOCC	-16,710	3312,000	-0,005	0,996
	TODCI	-16,380	2001,000	-0,008	0,993
	TOMET	-16,500	3590,000	-0,005	0,996
	TPAA	-0,745	1,694	-0,440	0,660

	TS	-16,630	3300,000	-0,005	0,996
Nota FAV		-1,083	0,468	-2,315	0,021 *