

iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Personalização da Experiência do Cliente nas Multi-Sided Platforms: Um Modelo Preditivo de Comportamento para Cross-selling no Setor Segurador

Rui Miguel Teixeira Costa

Mestrado em Gestão

Orientador:

Professor Doutor João Ricardo Paulo Marques Guerreiro,
Professor Associado com Agregação.

Iscte – Instituto Universitário de Lisboa

Outubro, 2025



DEPARTAMENTO DE MARKETING, OPERAÇÕES E GESTÃO GERAL

**Personalização da Experiência do Cliente nas Multi-Sided
Platforms: Um Modelo Preditivo de Comportamento para
Cross-selling no Setor Segurador**

Rui Miguel Teixeira Costa

Mestrado em Gestão

Orientador:

Professor Doutor João Ricardo Paulo Marques Guerreiro,
Professor Associado com Agregação.

Iscte – Instituto Universitário de Lisboa

Outubro, 2025

RESUMO

O propósito deste estudo é desenvolver um sistema preditivo baseado em *machine learning* para personalizar estratégias de *cross-selling* de seguros de vida na Doutor Finanças, MSP de referência no setor segurador português.

Implementou-se a metodologia *CRISP-DM* analisando a base de clientes reais através de variáveis comportamentais, transacionais e demográficas. Foram testados e avaliados quatro algoritmos de classificação supervisionada (*k-Nearest Neighbors*, *Logistic Regression*, *Naive Bayes* e *Decision Tree*) através das métricas CA e AUC, priorizando a capacidade de identificar oportunidades comerciais e minimizar oportunidades perdidas de conversão.

O modelo *Naive Bayes* demonstrou uma capacidade excepcional na identificação de clientes propensos à compra, enquanto o *Decision Tree* revelou perfis específicos de alta conversão: ausência de apólice automóvel anterior, origem em leads referenciadas e padrão de comunicação escrita normal. Os algoritmos identificaram segmentos de clientes com propensão superior de conversão, permitindo à empresa concentrar recursos nos perfis mais promissores e eliminar abordagens comerciais inadequadas.

A implementação transforma o *cross-selling* através da personalização da experiência do cliente. Esta abordagem substitui as campanhas genéricas por experiências comerciais personalizadas, assegurando a entrega de recomendações nos momentos certos do ciclo de vida familiar. Simultaneamente, reduz os custos operacionais, otimizando o contato com a base de clientes existente, e cria experiências diferenciadas que consolidam o relacionamento cliente-plataforma. A personalização da experiência do cliente, além de fortalecer o relacionamento, estabelece um ciclo virtuoso de aprendizagem contínua, no qual cada interação transforma-se numa oportunidade de refinamento que melhora continuamente a precisão preditiva.

Palavras-chave: Economia de Plataformas; Comportamento do Consumidor; Aprendizagem Automática; Estratégias de Marketing; Setor Segurador

Códigos JEL: G22, L14, D12, C45, M31

ABSTRACT

The purpose of this study is to develop a predictive system based on machine learning to personalise life insurance cross-selling strategies at Doutor Finanças, a leading MSP in the Portuguese insurance sector.

The CRISP-DM methodology was implemented by analysing the real client base through behavioural, transactional, and demographic variables. Four supervised classification algorithms (k-Nearest Neighbours, Logistic Regression, Naive Bayes, and Decision Tree) were tested and evaluated using CA and AUC metrics, prioritising the ability to identify genuine commercial opportunities and minimise missed conversion opportunities.

The Naive Bayes model demonstrated exceptional capability in identifying clients prone to purchase, whilst the Decision Tree revealed specific high-conversion profiles: absence of previous motor insurance policy, origin from referred leads, and normal written communication patterns. The algorithms identified client segments with superior conversion propensity, enabling the company to concentrate resources on the most promising profiles and eliminate inappropriate commercial approaches.

The implementation transforms cross-selling through the personalisation of the customer experience. This approach replaces generic campaigns with personalised commercial experiences, ensuring the delivery of recommendations at the right moments within the family lifecycle. Simultaneously, it reduces operational costs by optimising contact with the existing customer base, whilst creating differentiated experiences that consolidate the customer-platform relationship. Customer experience personalisation, beyond strengthening the relationship, establishes a virtuous cycle of continuous learning, wherein each interaction transforms into a refinement opportunity that continuously enhances predictive accuracy.

Keywords: Platform Economics; Consumer Behavior; Machine Learning; Marketing Strategies; Insurance Industry

JEL Codes: G22, L14, D12, C45, M31

ÍNDICE

RESUMO	i
ABSTRACT	iii
1. INTRODUÇÃO	1
2. REVISÃO DA LITERATURA	7
2.1. <i>MARKETING RELACIONAL E GESTÃO DO RELACIONAMENTO COM O CLIENTE</i>	7
2.1.1. Fundamentos do marketing relacional	7
2.1.2. Compromisso e confiança como variáveis mediadoras	7
2.1.3. Processo de satisfação do cliente	8
2.1.4. Fundamentos estratégicos e organizacionais da GRC	9
2.1.5. Aplicação estratégica da GRC nas MSPs no setor segurador	10
2.1.6. Síntese crítica e convergências paradigmáticas	11
2.2. <i>PERSONALIZAÇÃO DA EXPERIÊNCIA DO CLIENTE E IA</i>	12
2.2.1. Evolução conceptual da experiência do cliente	12
2.2.2. Fragmentação das medidas de experiência do cliente	14
2.2.3. Personalização: Dimensões multidimensionais e paradoxos	15
2.2.4. A IA como <i>enabler</i> da personalização	17
2.2.5. Síntese crítica e convergências paradigmáticas	18
2.3. <i>MULTI-SIDED PLATFORMS NO SETOR SEGURADOR</i>	20
2.3.1. Definição e arquitetura das MSPs	20
2.3.2. Transformação da percepção da criação de valor	20
2.3.3. Importância das MSPs no setor segurador	21
2.3.4. Síntese crítica e convergências paradigmáticas	22
2.4. <i>CROSS-SELLING E ALGORITMOS PREDITIVOS NO SETOR SEGURADOR</i>	24
2.4.1. <i>Cross-selling</i> : Fundamentos conceptuais e estratégicos	24
2.4.2. Tensões teóricas e transformação paradigmática	24
2.4.3. Influência da personalização no <i>cross-selling</i>	25
2.4.4. <i>Machine learning</i> : algoritmos preditivos supervisionados	26
2.4.5. Síntese crítica e convergências paradigmáticas	27
3. METODOLOGIA	29
3.1. <i>COMPREENSÃO DO NEGÓCIO (BUSINESS UNDERSTANDING)</i>	29
3.1.1. Framework analítico das msp: dimensões estratégicas no setor segurador	29

3.1.2. Cocriação de valor e efeitos de rede	31
3.1.3. Contexto estratégico e oportunidade de mercado	33
3.1.4. Objetivos estratégicos para o negócio da modelação preditiva	34
3.2. COMPREENSÃO DOS DADOS (DATA UNDERSTANDING)	35
3.2.1. Caracterização da amostra e estrutura do <i>dataset</i>	35
3.2.2. Dimensão comportamental e perfil sociodemográfico	36
3.2.3. Padrões comportamentais	37
3.3. PREPARAÇÃO DOS DADOS (DATA PREPARATION)	37
3.3.1. Estratégia de tratamento das variáveis	37
3.3.2. Caracterização estrutural do <i>dataset</i>	38
3.3.3. Avaliação da qualidade dos dados	38
3.3.4. Identificação e análise de limitações	38
3.3.5. Implementação do tratamento de dados	38
3.4. MODELAÇÃO (MODELING)	39
3.4.1. Plataforma de implementação	39
3.4.2. Arquitetura do workflow de dados	39
3.4.3. Estratégia de modelação: algoritmos para cross-selling	39
3.4.4. Métricas de avaliação	40
3.4.5. Matriz de confusão	41
3.4.6. Framework de treino e testes	42
3.5. AVALIAÇÃO (EVALUATION)	44
3.5.1. Introdução sintética	44
3.5.2. Análise de performance global	44
3.5.3. Análise de <i>overfitting</i>	44
3.5.4. Perfis comportamentais dos modelos	46
3.5.5. Análise granular das matrizes de confusão	47
3.6. CONCLUSÕES (DEPLOYMENT)	49
3.6.1. Validação empírica e síntese dos resultados	49
3.6.2. Descoberta de padrões: Algoritmo <i>Tree</i>	50
3.6.1. Seleção do modelo ótimo	52
4. CONSIDERAÇÕES FINAIS / DISCUSSÃO	53
5. RECOMENDAÇÕES PARA INVESTIGAÇÕES FUTURAS	55
6. LIMITAÇÕES	57
7. REFERÊNCIAS BIBLIOGRÁFICAS	59

Índice de imagens

Imagem 1 (origem Orange): Matriz confusão - <i>Logistic Regression</i>	78
Imagem 2 (origem Orange): Matriz confusão - <i>kNN</i>	78
Imagem 3 (origem Orange): Matriz confusão - <i>Naive Bayes</i>	78
Imagem 4 (origem Orange): Matriz confusão - <i>Tree</i>	78
Imagem 5 (origem Orange): Nó terminal da <i>decison tree</i> - Árvore para “PossuiApoliceVida2024:Sim”	79
Imagem 6 (origem Orange): Nó terminal da <i>decison tree</i> – Árvore para “PossuiApoliceVida2024:Não”	79

Índice de tabelas

Tabela 1: Performance em Treino	73
Tabela 2: Performance em Teste real	73
Tabela 3: Comparação - Treino vs Teste.....	73
Tabela 4: Ranking Geral (AUC + CA)	73
Tabela 5: Segmentação dos rendimentos	74
Tabela 6: Segmentação Geracional por Faixas Etárias	74
Tabela 7: Variáveis de comportamento relacional	74
Tabela 8: Estrutura de variáveis do <i>dataset</i>	75

Glossário de abreviaturas

ASF - Autoridade De Supervisão De Seguros E Fundos De Pensões

PP - Pontos Percentuais

EU - União Europeia

PIB - Produto Interno Bruto

MSP - Multi-Sided Platform

IA - Inteligência Artificial

ML - Machine Learning

CA - Accuracy

AUC - Área Sob A Curva

SVM - Support Vector Machine

CRISP-DM – Cross Industry Standard Process For Data Mining

CRM - Customer Relationship Management

GRC - Gestão Do Relacionamento Com O Cliente

CLV - Customer Lifetime Value

MPE - Modelo De Probabilidade De Elaboração

TUG - Teoria Dos Usos E Gratificações

TCS -Teoria Cognitiva Social

RGPD – Regulamento Geral Sobre A Proteção De Dados

CPCC - Crédito Pessoal E Crédito Consolidado

CH - Crédito Habitação

1. INTRODUÇÃO

O setor segurador mundial tem demonstrado uma trajetória de crescimento notável, apresentando em 2023 um crescimento robusto de 6.1%, superando significativamente o crescimento do PIB global (2.7%), o que evidencia a expansão da procura por produtos de proteção financeira (Autoridade de Supervisão de Seguros e Fundos de Pensões, 2024). Com um volume total de 7.186 mil milhões de dólares americanos, o mercado consolidou a sua trajetória ascendente, contrastando com a desaceleração económica mundial face aos 3,1% registados em 2022 (Autoridade de Supervisão de Seguros e Fundos de Pensões, 2024). Esta performance reflete não apenas a resiliência do setor, mas também a crescente consciencialização sobre a importância da proteção financeira num contexto de incerteza económica global.

A distribuição geográfica do mercado segurador revela uma hegemonia norte-americana consolidada, com 47% da quota mundial, registando um incremento de 0.7 pontos percentuais (pp). A região asiática, apesar de manter uma posição relevante com 24% do mercado, evidenciou uma ligeira contração de -1.1 pp, enquanto a Europa manteve estável os seus 23% de mercado (Autoridade de Supervisão de Seguros e Fundos de Pensões, 2024). Esta configuração geográfica sublinha a importância dos mercados desenvolvidos na dinâmica global do setor segurador.

A nível mundial, o ramo não vida demonstrou um crescimento robusto de 7.6%, praticamente duplicando o desempenho de 2022, refletindo a crescente consciencialização sobre riscos patrimoniais e de responsabilidade civil. O ramo vida, após uma contração (-4.3%) em 2022, registou uma recuperação significativa (+3.9%), indicando uma normalização das condições de mercado (Autoridade de Supervisão de Seguros e Fundos de Pensões, 2024). Esta divergência de comportamentos entre ramos demonstra a complexidade das dinâmicas do mercado e a necessidade de estratégias diferenciadas entre ramos.

O mercado segurador na União Europeia (EU) superou o desempenho global, registando um crescimento nominal de 6.9% e atingindo 1.2 biliões de dólares. Esta performance foi impulsionada primordialmente pelo ramo não vida (+9.7%), que beneficiou da pressão inflacionista sobre os prémios e pela crescente procura por coberturas especializadas. O ramo vida, embora com crescimento mais moderado (+4%), demonstrou sinais de estabilização após alguns períodos de volatilidade (Autoridade de Supervisão de Seguros e Fundos de Pensões, 2024).

Contrastando com esta tendência positiva da UE, o mercado segurador português apresentou uma performance divergente, com uma produção total de 10,9 mil milhões de euros, registando uma ligeira contração (-1%) face a 2022. Esta evolução reflete as especificidades estruturais do mercado segurador nacional, em que este representa 4.1% do PIB nacional e uma quota estável de aproximadamente 1.1% no mercado europeu, posicionando-se no 14º lugar do ranking continental (Autoridade de Supervisão de Seguros e Fundos de Pensões, 2024).

Numa análise mais detalhada ao mercado português, em 2023, o ramo vida contraiu significativamente (-12.4%), enquanto o ramo não vida evidenciou um crescimento sustentado (+11.1%). No contexto nacional, estas divergências refletem os baixos níveis de poupança e a necessidade premente de inovação nos produtos de seguro de vida, contrastando com a dinâmica positiva que já acontece nos seguros do ramo não vida. A estrutura concorrencial em Portugal é muito concentrada, com as cinco principais seguradoras a deterem 59.5% do mercado global, lideradas pela Fidelidade (29.5%). No que se refere ao ramo não vida, a concentração é ainda mais acentuada, com o top 5 a representar 74% do mercado (Autoridade de Supervisão de Seguros e Fundos de Pensões, 2024).

O mercado segurador português enfrenta desafios estruturais que exigem estratégias diferenciadas. A pressão inflacionista sobre os custos de sinistros e os reduzidos níveis de poupança nacional constituem constrangimentos significativos, exigindo inovação e adaptação às necessidades emergentes dos consumidores para manter a competitividade no contexto europeu. É possível concluir que a inovação é fundamental para o mercado crescer de forma consistente, principalmente no ramo vida (Autoridade de Supervisão de Seguros e Fundos de Pensões, 2024).

A inovação deve abranger múltiplas dimensões, incluindo o desenvolvimento de produtos mais flexíveis e adaptados às necessidades específicas dos consumidores portugueses, a melhoria dos canais de distribuição e comunicação, bem como a otimização das estruturas de custos para oferecer propostas de valor mais competitivas. A melhoria dos canais de distribuição e comunicação são um ponto crítico e que exigem muito investimento, principalmente para aumentar a literacia financeira dos consumidores e com isso a consciencialização sobre a importância dos seguros, nomeadamente dos seguros de vida (Autoridade de Supervisão de Seguros e Fundos de Pensões, 2024).

Neste contexto de transformação, as *Multi-Sided Platforms* (MSPs) emergem como atores disruptivos no setor segurador, alavancando o seu acesso a uma vasta quantidade de dados de clientes, que são utilizados para influenciar a jornada de tomada de decisões, a gestão personalizada do relacionamento com o cliente, bem como a retenção e fidelização (Belhadi et al., 2023; Ijomah et al., 2024; V. Kumar et al., 2019; Pousttchi & Gleiss, 2019). Estas plataformas têm um impacto significativo no setor segurador, mudando a perceção de como o valor é criado, com especial ênfase na jornada de compra de seguros, potenciando as tecnologias digitais para oferecer uma proposta de valor centrada nos principais fatores de decisão de compra do cliente (Pousttchi & Gleiss, 2019).

Para as MSPs alcançarem uma experiência de cliente cada vez mais impactante, é essencial usarem técnicas de inteligência artificial (IA) como *machine learning* (ML) para criar modelos preditivos que antecipem as necessidades e preferências dos clientes (Orenga-Roglá & Chalmeta, 2016). Com a utilização de modelos preditivos é possível ambicionar ter uma gestão personalizada do relacionamento com o cliente, para criar experiências personalizadas cada vez mais envolventes, bem como prever as

necessidades futuras do cliente e disponibilizar recomendações personalizadas, aumentando a satisfação e a fidelidade do cliente, bem como o volume de vendas (Ijomah et al., 2024; V. Kumar et al., 2019).

A assimetria do acesso à informação, entre as MSPs e as seguradoras, constitui um ativo estratégico de elevada relevância, conferindo às MSPs uma visão transversal e integrada do comportamento dos consumidores no setor segurador que transcende os silos informacionais característicos das seguradoras, que têm acesso apenas a dados dos seus clientes e estão mais focadas em usar a IA para classificar o risco associado ao perfil de cliente (fraude) e probabilidade de sinistralidade (Ngai et al., 2011). A amplitude e heterogeneidade dos dados a que as MSPs têm acesso, possibilitam uma compreensão holística das necessidades e preferências dos clientes (McFall et al., 2020; Pousttchi & Gleiss, 2019), uma vez que possuem uma diversificação das suas carteiras de seguros por múltiplas seguradoras, em que o mesmo cliente, pode ter o seguros em diferentes seguradoras. Este posicionamento privilegiado na arquitetura de acesso à informação representa uma vantagem competitiva muito importante, no setor, e é igualmente um potencial analítico diferenciador que, quando devidamente explorado através de metodologias científicas adequadas, poderá resultar em inferências de valor estratégico, inacessível ao mercado de mediação tradicional, bem como às próprias seguradoras, por não terem acesso à informação transversal sobre os hábitos e comportamentos de compra dos clientes (Belhadi et al., 2023; Ijomah et al., 2024).

A decisão de compra de seguros de vida por parte dos consumidores é influenciada por diversos fatores, entre eles estão fatores relacionados com o indivíduo, como estado de saúde, nível de literacia financeira (Bhatia et al., 2024), ocupação, fatores relacionados com a família, como número de filhos, fatores financeiros como rendimento, fatores comportamentais como aversão ao risco, deixar herança para os filhos, vieses comportamentais como aversão à perda, experiência de morte familiar, fatores associados ao produto, fatores de serviço como a confiança na seguradora e nos distribuidores, a qualidade do serviço prestado, a experiência passada com outros seguros, entre outros (Bhatia et al., 2021; Oluwaseyitan et al., 2023; A. K. Srivastava & Jadid, 2020).

Um baixo nível de confiança nas seguradoras e distribuidores, por parte dos consumidores, desempenha um papel fundamental na formação da perceção do seguro como um custo, fomentando um afastamento dos indivíduos na procura ativa por mais conhecimento sobre os seguros, o que acaba por ser um fator importante para que os consumidores não tenham um nível elevado de literacia financeira (Fisch & Seligman, 2022). Indivíduos com um nível elevado de literacia financeira estão mais inclinados a comprar um seguro de vida, o que sugere que a compreensão dos conceitos financeiros desempenha um papel importante no processo de decisão de compra (C. Lin et al., 2017; Weedige et al., 2019).

No contexto da transformação digital contemporânea e da crescente importância estratégica da personalização, a capacidade da IA para processar e interpretar grandes volumes de dados com objetivos

específicos tem sido amplamente explorada para criar jornadas de compra personalizadas (V. Kumar et al., 2019). As empresas para conseguirem ter experiências cada vez mais impactantes, têm de usar técnicas de IA como o ML para criar modelos preditivos que antecipem algum dos comportamentos do cliente (Orenga-Roglá & Chalmeta, 2016). A personalização emerge assim, como o elemento diferenciador fundamental (Chandra et al., 2022), numa estratégia focada no cliente (Payne & Frow, 2006), que evoluiu de um benefício adicional para um requisito essencial na experiência do cliente na sua jornada de compra (Weidig et al., 2024), permitindo às empresas conseguirem obter um maior nível de *engagement*, de retenção e fidelização, um maior grau de satisfação, bem como aumentar o volume de vendas (Gupta et al., 2018; Orenga-Roglá & Chalmeta, 2016).

No presente contexto, a empresa Doutor Finanças que atua como uma *Multi-sided Platform* (MSP), nativa digital, pretende revolucionar a forma como as pessoas percecionam os seguros, transformando-os num fator fundamental para o seu bem-estar financeiro. Para a Doutor Finanças, a atividade de mediação de seguros tem um propósito muito claro e transformador de “*cuidar e proteger o bem-estar financeiro da sociedade*”. Este propósito está alicerçado num pilar muito forte de literacia financeira e, ao aumentar o nível geral de literacia financeira da sociedade (Bhatia et al., 2024; Fisch & Seligman, 2022), é possível transformar a perceção que os indivíduos têm sobre os seguros, principalmente os seguros de vida (Bhatia et al., 2021; C. Lin et al., 2017), elevando-os de um simples custo necessário para um investimento fundamental na proteção e bem-estar financeiro. Como MSP no mercado segurador, a Doutor Finanças, sempre esteve muito focado na gestão do relacionamento com o cliente (Catlin et al., 2018) e na recolha dos dados nos diferentes pontos de contacto com o cliente (Weidig et al., 2024), porque esses dados são fundamentais para criar uma experiência de cliente personalizada (Rusnaini et al., 2024).

Considerando esta perspetiva, para a Doutor Finanças a personalização da experiência do cliente é vista como um fator crítico (Weidig et al., 2024) para aumentar o envolvimento comercial do cliente isto é, conseguir vender produtos adicionais a cada cliente (Ghoshal et al., 2021), captar a sua atenção (Tolga Akçura & Srinivasan, 2005), ganhar a sua confiança (Aka et al., 2016), criar vínculos emocionais (Weidig et al., 2024), assegurar que os clientes consigam obter uma experiência positiva antes da compra, durante a compra e no pós compra de seguros (Lemon & Verhoef, 2016; Verhoef et al., 2009). Esta personalização da experiência do cliente vai permitir à Doutor Finanças conseguir obter um maior nível de *engagement*, satisfação, retenção e fidelização, bem como aumentar o volume de vendas (Gupta et al., 2018; Orenga-Roglá & Chalmeta, 2016).

Este projeto vai investigar a personalização da experiência do cliente, de uma MSP, com a utilização de técnicas de IA, mais especificamente de ML, para identificar os algoritmos preditivos de comportamento com melhores resultados para apoiar o *cross-selling* de seguros de vida, a clientes da carteira de seguros da Doutor Finanças. Espera-se que os resultados proporcionem *insights* significativos, com particular relevância para a transformação digital e inovação no setor segurador, para

a otimização de estratégias de personalização da experiência do cliente, em contextos de MSP, contribuindo para a maximização do valor do ciclo de vida do cliente através de abordagens de *cross-selling* personalizadas, fomentando a venda de seguros de vida, à atual carteira de clientes.

Apesar do crescente interesse acadêmico na aplicação de técnicas de IA no setor segurador, a literatura científica apresenta lacunas significativas na intersecção entre MSPs, personalização da experiência do cliente e *cross-selling* de seguros de vida. A maioria dos estudos existentes foca-se isoladamente em: (i) aplicações de ML para classificar o risco associado ao perfil de cliente, detetar fraude e probabilidade de sinistralidade, por parte das seguradoras, (Aldoseri et al., 2024; Ngai et al., 2011) e (ii) fatores comportamentais na decisão de compra de seguros (Oluwaseyitan et al., 2023).

Os estudos académicos relacionados com seguros concentram-se predominantemente na customização de produtos, risco de sinistralidade e risco associado ao cliente na contratação do seguro, existindo muitas lacunas sobre a experiência do cliente na sua jornada de compra de seguros (McFall et al., 2020). Contudo, permanece por explorar a aplicação sistemática de modelos preditivos de ML especificamente direcionados para o *cross-selling* de seguros de vida em contextos de MSP, particularmente considerando as especificidades do mercado segurador português e as vantagens informacionais únicas destas plataformas. Esta lacuna é particularmente relevante dado que as MSPs possuem acesso privilegiado a dados transversais sobre comportamentos de compra que transcendem os silos informacionais das seguradoras tradicionais.

A presente investigação pretende colmatar esta lacuna mediante contribuições científicas na análise comparativa sistemática do desempenho de diferentes algoritmos de ML direcionados para o *cross-selling* de seguros de vida, aplicados a dados reais de uma MSP portuguesa. O estudo fornece evidência empírica sobre a eficácia relativa das diferentes abordagens preditivas neste contexto específico, contribuindo para a criação de estratégias de personalização da experiência do cliente baseadas em evidências científicas. Os resultados têm aplicabilidade direta para MSPs no setor segurador, contribuindo para a transformação digital e inovação do setor através de metodologias cientificamente validadas.

A presente investigação tem como objetivo central a personalização da experiência do cliente das MSPs, visando otimizar estratégias de *cross-selling* no setor segurador através do desenvolvimento e validação de modelos preditivos do comportamento do consumidor, utilizando uma amostra representativa da base de dados de clientes da carteira de seguros da Doutor Finanças. No contexto contemporâneo da economia digital, onde as MSPs emergem como ecossistemas complexos de criação de valor derivado da capacidade de recolher dados, da conectividade e dos efeitos de rede (Mlčúchová, 2022), o objetivo principal do projeto consiste na estimação probabilística da propensão de aquisição de seguros de vida por parte de clientes detentores de outras apólices, permitindo a conceção de estratégias

de *cross-selling* personalizadas para aperfeiçoar a experiência do cliente ao longo da sua jornada de compra (Li et al., 2011).

Esta investigação representa uma contribuição original para o corpus científico em três dimensões fundamentais. Dimensão teórica: em que contribui para a teoria da personalização nas MSPs, no setor segurador, através da integração de modelos comportamentais do consumidor com técnicas avançadas de ML, expandindo o conhecimento sobre a aplicação de IA em contextos de intermediação financeira. Dimensão prática: na análise do desempenho dos modelos preditivos de *cross-selling* nas MSPs, estabelecendo *benchmarks* de desempenho específicos para o setor segurador que poderão ser adotados por investigações futuras. Dimensão de mercado: em que fornece *insights* acionáveis para o setor segurador português, particularmente relevantes no contexto de transformação digital e necessidade de inovação identificada pela ASF (2024), contribuindo para a competitividade do setor nacional no panorama europeu.

A originalidade desta investigação reside na convergência inédita entre: (i) dados proprietários de uma MSP portuguesa, previamente submetidos a processo de anonimização; (ii) aplicação específica ao *cross-selling* de seguros de vida; e (iii) criar *benchmarks* de desempenho específicos de algoritmos de ML no setor segurador nacional, constituindo um contributo singular para a literatura científica na intersecção entre *insurtechs*, e ML.

2. REVISÃO DA LITERATURA

2.1. MARKETING RELACIONAL E GESTÃO DO RELACIONAMENTO COM O CLIENTE

2.1.1. FUNDAMENTOS DO MARKETING RELACIONAL

O marketing relacional (L. Berry, 2002) surge como uma filosofia estratégica orientada para criar, manter e melhorar relacionamentos fortes com clientes e outros *stakeholders* dentro e fora das empresas (Rosário & Casaca, 2023). Esta abordagem representa uma transição paradigmática fundamental na orientação empresarial, transitando do foco da atração de clientes para a construção e manutenção de relacionamentos de longo prazo com os clientes como melhor forma de manter a competitividade (Palmatier et al., 2006).

A distinção entre marketing transacional e relacional é fundamental para compreender esta evolução. O marketing transacional foca na atração de clientes através de transações individuais, com orientação para vendas únicas e uma filosofia de troca de valor, sendo tipicamente responsabilidade apenas do departamento de marketing. Por outro lado, o marketing relacional concentra-se na construção de relacionamentos de longo prazo, priorizando múltiplas interações, construção de valor mútuo e envolvendo toda a organização no processo (Aka et al., 2016; Kozlenkova et al., 2014).

A evolução do marketing relacional pode ser compreendida através de quatro estágios históricos distintos. O marketing relacional inicial (1950s-1980s) focou-se na construção de relacionamentos de longo prazo com clientes, reconhecendo a importância do negócio repetido e da fidelidade através de métodos principalmente *offline* como correio direto, telemarketing e vendas diretas (Rosário & Casaca, 2023). O marketing de base de dados (1980s-1990s) foi impulsionado pelo advento de computadores e bases de dados, permitindo a recolha e armazenamento mais eficaz de informação dos clientes, facilitando a mudança do *mass marketing* (Borden, 1964) para o marketing um-para-um (Peppers & Rogers, 2016) através de interações individualizadas.

O terceiro estágio, *customer relationship management (CRM)* (1990s-2000s), caracterizou-se pela introdução de sistemas para gestão de informação, *tracking* de interações com os clientes e automação de campanhas, com ênfase na melhoria do serviço ao cliente, compreensão das necessidades e experiências personalizadas, período em que os programas de fidelização e retenção ganharam proeminência. Por último, a era Mobile e Social Media (2010s-presente) caracteriza-se pela proliferação de smartphones e plataformas de redes sociais. Esta época criou o desafio da gestão de múltiplos *touchpoints* digitais. Simultaneamente, incorporou elementos essenciais como o *engagement* em tempo real, feedback instantâneo e o marketing de influenciadores (Leung et al., 2022; Rosário & Casaca, 2023).

2.1.2. COMPROMISSO E CONFIANÇA COMO VARIÁVEIS MEDIADORAS

O compromisso e a confiança emergem como variáveis mediadoras-chave no modelo de marketing relacional, distinguindo-se claramente das abordagens tradicionais.

Enquanto o marketing tradicional foca-se na aquisição de novos clientes através de "*offensive marketing*" (Davidson & O'Reilly, 1987), o marketing relacional adota uma estratégia defensiva com duas abordagens complementares: (i) aquisição de clientes, reconhecendo a necessidade contínua de captar novos clientes para o negócio e as implicações na rentabilidade com a queda no número total de clientes, especialmente face ao desafio crescente com a entrada de produtos substitutos e novos concorrentes; e (ii) retenção de clientes, priorizando a otimização da experiência do cliente e níveis de satisfação, constituindo a convicção principal do marketing relacional com prioridade no marketing de retenção primeiro, e aquisição, segundo (Aka et al., 2016).

Esta abordagem estrutura-se em cinco componentes estratégicos fundamentais: (i) *Core Service Strategy* (Kasabov, 2015), que serve como base para venda de serviços adicionais ao longo do tempo; (ii) *Relationship Customization* (L. Berry, 2002), focada na personalização de relacionamentos; (iii) *Service Augmentation* (Levitt, 1980), almejando o aumento e melhoria dos serviços; (iv) *Relationship Pricing* (Bolton & Lemon, 1999), estabelecendo preços baseados no relacionamento com o cliente; e (v) *Internal Market* (L. L. Berry, 1995), objetivando desenvolver o mercado interno da organização.

O marketing relacional fundamenta-se em quatro determinantes estratégicos interdependentes que configuram a arquitetura das relações comerciais sustentáveis (Palmatier et al., 2018). (i) A confiança emerge como constructo psicológico basilar, caracterizada pela predisposição cognitiva e emocional do cliente para depositar expectativas positivas nas competências e integridade do fornecedor (Fournier, 1998). Materializando-se através do cumprimento sistemático dos compromissos contratuais e promessas comerciais, o que resulta na consolidação de vínculos relacionais duradouros e na otimização dos índices de satisfação; (ii) o compromisso constitui a dimensão volitiva e estratégica da díade comercial, representando a intenção deliberada e o investimento mútuo de recursos entre vendedor e comprador na construção e manutenção das parcerias comerciais economicamente viáveis e mutuamente benéficas a longo prazo, contribuindo para a edificação de vínculos relacionais simbióticos e para a consolidação de uma imagem corporativa diferenciada; (iii) a comunicação configura-se como processo bidirecional e personalizado de troca informacional, caracterizado pela individualização das mensagens dirigidas a cada cliente, integrando de forma sinérgica todos os *touchpoints* comunicacionais da estratégia de marketing com o objetivo de assegurar consistência na promessa de valor e coerência na identidade corporativa transmitida; (iv) a qualidade do serviço posiciona-se como elemento diferenciador crítico na obtenção de vantagem competitiva sustentável, operacionalizando-se através da entrega consistente de experiências de serviço que excedem as expectativas dos clientes, resultando na maximização dos níveis de satisfação, *engagement* e na consolidação da fidelização do cliente (Kwak et al., 2019; Palmatier et al., 2006, 2018).

2.1.3. PROCESSO DE SATISFAÇÃO DO CLIENTE

A satisfação do cliente constitui um constructo multidimensional (Oliver, 2010), definido como, a medida avaliativa de como os produtos ou serviços fornecidos atendem ou superam as expectativas

previamente formadas pelo cliente. Este processo cognitivo-emocional desenvolve-se através de uma sequência estruturada de etapas interdependentes: (i) formação de expectativas pelo cliente baseada em experiências anteriores e informação disponível; (ii) experiência efetiva de consumo do produto ou serviço; (iii) processo comparativo entre a experiência vivenciada e as expectativas iniciais; e (iv) avaliação resultante que culmina em satisfação (quando a performance supera ou iguala expectativas) ou insatisfação (quando a performance fica aquém das expectativas) (Aka et al., 2016).

A relação intrínseca entre satisfação e percepção de valor emerge como elemento fundamental na dinâmica comportamental do consumidor, uma vez que os clientes procuram maximizar o valor obtido e formam expectativas correspondentes a essa otimização (Zeithaml, 1988). Quando a oferta não corresponde ao valor esperado, verifica-se um impacto negativo direto na satisfação do cliente, com consequências subsequentes na intenção de recompra e nos níveis de retenção e fidelização do cliente. Consequentemente, a abordagem estratégica de retenção fundamenta-se na entrega consistente de elevados níveis de satisfação, resultando na consolidação de vínculos relacionais duradouros e no fortalecimento da fidelização do cliente (Huang et al., 2019).

O processo de satisfação segue uma arquitetura estruturada que conduz ao resultado final desejado, operacionalizando-se através das seguintes fases sequenciais: recolha sistemática de informação sobre clientes mediante múltiplas interações e *touchpoints*; análise e interpretação aprofundada das necessidades, preferências e expectativas identificadas; integração estratégica dos *insights* obtidos nas estratégias de marketing e desenvolvimento de produtos; entrega personalizada e diferenciada de produtos e serviços alinhados com as expectativas do cliente; processo comparativo entre a performance entregue e as expectativas iniciais; culminando na formação de sentimentos positivos que se traduzem na satisfação efetiva do cliente (L. Berry, 2002; Olufunke Anne Alabi et al., 2024).

Os fatores críticos de sucesso que determinam a eficácia deste processo incluem: a disponibilização de informação suficiente e transparente sobre produtos e serviços oferecidos, a consideração sistemática do feedback do cliente nos processos de desenvolvimento e de melhoria contínua, a implementação de uma comunicação eficaz, transparente e bidirecional, a personalização estratégica de ofertas e experiências de acordo com perfis individuais, e por fim, a construção progressiva de confiança através de ações consistentes e cumprimento sistemático de promessas (Rosário & Casaca, 2023).

2.1.4. FUNDAMENTOS ESTRATÉGICOS E ORGANIZACIONAIS DA GRC

A gestão do relacionamento com o cliente (GRC) representa um paradigma estratégico integrado que transcende amplamente a mera implementação tecnológica, exigindo uma transformação organizacional profunda orientada para uma visão *customer-centric* (Levitt, 1960). Tal abordagem requer a eliminação de silos departamentais e a harmonização dos objetivos organizacionais com estratégias centradas no cliente, fundamentando-se na cocriação de valor entre empresa e consumidor (Payne & Frow, 2006).

A implementação eficaz de estratégias GRC enfrenta desafios multidimensionais que se articulam em três domínios críticos, são eles, tecnológicos (integração de sistemas, gestão de *big data*, seleção de soluções adequadas), organizacionais (resistência à mudança, desenvolvimento de competências) e estratégicos (alinhamento de objetivos, definição de métricas, gestão de expectativas) (Karthick et al., 2023). A referida complexidade multifacetada exige uma abordagem holística que reconheça as interdependências sistêmicas entre os diversos componentes organizacionais.

Contemporaneamente, a GRC experimenta uma transformação paradigmática fundamental, caracterizada pela transição de abordagens transacionais para estratégias centradas no *engagement* (Hollebeek et al., 2014) e cocriação de valor (Orenga-Roglá & Chalmeta, 2016). Esta evolução reflete uma mudança cultural profunda nas expectativas dos consumidores, que procuram uma participação ativa em todas as fases do desenvolvimento de produtos e desejam partilhar experiências através das suas redes de contacto (Sheth et al., 2023).

Os fatores críticos de sucesso de uma estratégia de GRC estruturam-se hierarquicamente numa arquitetura complexa. Fatores organizacionais (liderança, recursos humanos qualificados, estrutura organizacional) e experiência temporal impactam diretamente o sucesso da estratégia de GRC. Por contraste, a gestão de conhecimento, a tecnologia e a orientação *customer-centric* exercem influência indireta por intermédio dos fatores organizacionais (Garrido-Moreno & Padilla-Meléndez, 2011). Tal hierarquização evidencia a primazia dos elementos organizacionais como determinantes do sucesso estratégico, estabelecendo as bases conceptuais para a aplicação prática em contextos específicos de mercado (Karthick et al., 2023).

Neste contexto evolutivo, a personalização emerge como elemento diferenciador crítico, transcendendo a sua conceção tradicional de benefício adicional para tornar-se requisito essencial na prestação de serviços contemporâneos. A operacionalização materializa-se através da adaptação estratégica dos pontos de contacto às necessidades individuais dos clientes, resultando em níveis superiores de *engagement*, satisfação, retenção, fidelização e performance comercial (Orenga-Roglá & Chalmeta, 2016). Para alcançar experiências verdadeiramente impactantes, as organizações implementam técnicas de IA, nomeadamente ML, desenvolvendo modelos preditivos para prever os comportamentos dos clientes (Roy et al., 2025).

2.1.5. APLICAÇÃO ESTRATÉGICA DA GRC NAS MSPS NO SETOR SEGURADOR

A capacidade organizacional de melhorar a gestão do relacionamento do cliente correlaciona-se diretamente com a retenção e fidelização de clientes, representando um determinante fundamental do *Customer Lifetime Value* (CLV) (Winer, 2001). A convergência tecnológica proporcionou uma revolução nas formas de interação, em que a internet conferiu flexibilidade interacional sem precedentes, enquanto a IA dotou as organizações de capacidade de resposta em tempo real (Babatunde et al., 2024).

Tal convergência permite a transição estratégica de abordagens centradas na aquisição para paradigmas focalizados na retenção (Aka et al., 2016).

A otimização da jornada do cliente resulta numa sequência de benefícios organizacionais através de um processo sequencial com um *engagement* superior e satisfação elevada (Hollebeek et al., 2016), resulta numa maior retenção e fidelização, traduzindo-se num CLV superior e receitas incrementais sustentáveis (Winer, 2001). Este impacto revela-se particularmente transformador para pequenas e médias empresas (PMEs), como mediadoras de seguros, uma vez que Nnenna Ijeoma Okeke et al., (2024) demonstram que as PMEs conseguem agora competir com grandes corporações através de motores de recomendação inteligentes que potenciam *upselling* e *cross-selling* enquanto reduzem perda de clientes, resultando em maior CLV. Para monitorizar tal performance, as organizações contemporâneas utilizam métricas centradas no cliente, designadamente custos de aquisição, taxas de conversão, índices de retenção/*churn*, vendas recorrentes e *customer share* (Xing et al., 2023).

No setor segurador, a dinâmica económica evidencia uma assimetria fundamental manifestada pelo facto do custo de aquisição de novos clientes exceder em 7 a 9 vezes o custo de retenção, tornando imperativa a implementação de estratégias de fidelização (Ortega et al., 2023). Tal disparidade económica adquire particular relevância num contexto onde as seguradoras tradicionais enfrentam desafios sistémicos de confiança e experiência do cliente (Ghoshal et al., 2021), configurando simultaneamente oportunidades estratégicas para as Multi-Sided Platforms como intermediários digitais.

As MSPs posicionam-se estrategicamente mediante quatro pilares fundamentais que englobam: transparência operacional, confiança, informação estruturada e simplicidade processual. Através desta arquitetura conceptual, mediam eficazmente a relação seguradora-cliente e conferem credibilidade ao processo transacional. O acesso privilegiado a dados comportamentais constitui, por conseguinte, um ativo estratégico que permite às MSPs desenvolver modelos preditivos sofisticados, resultando em maior *engagement*, retenção superior e uma melhor performance comercial (Ortega et al., 2023).

2.1.6. SÍNTESE CRÍTICA E CONVERGÊNCIAS PARADIGMÁTICAS

A implementação de uma visão *customer-centric* representa um determinante estratégico da satisfação e lealdade dos clientes, refletindo a capacidade organizacional de criar valor e relacionamentos duradouros (Ramaj & Ismaili, 2015). Semelhante abordagem exige compreensão profunda e multidimensional das necessidades, hábitos e comportamentos de compra (Ramaj & Ismaili, 2015). A transformação de foco transacional para relacional caracteriza as organizações (Aka et al., 2016; Rosário & Casaca, 2023) que equilibram estrategicamente abordagens de aquisição e retenção, configurando um modelo híbrido de gestão (Ramaj & Ismaili, 2015).

A convergência entre GRC, modelos preditivos e MSPs manifesta-se como um ecossistema estratégico no setor segurador. A capacidade de antecipar comportamentos através de modelos

preditivos sofisticados, combinada com experiências em tempo real, representa vantagem competitiva sustentável que transcende a mera diferenciação (Sadiq et al., 2025). Tal convergência tecnológico-estratégica materializa-se não apenas como evolução incremental, mas como transformação paradigmática que redefine os fundamentos da competitividade no setor segurador contemporâneo, estabelecendo novos padrões de excelência na gestão do relacionamento com o cliente e criação de valor sustentável. Esta criação de valor sustentável materializa-se, na prática, através da personalização sistemática da experiência do cliente, exigindo uma análise aprofundada dos seus fundamentos conceptuais e das tecnologias que a viabilizam (Akhavan & Hassannayebi, 2024).

2.2. PERSONALIZAÇÃO DA EXPERIÊNCIA DO CLIENTE E IA

2.2.1. EVOLUÇÃO CONCEPTUAL DA EXPERIÊNCIA DO CLIENTE

A experiência do cliente constitui um constructo multidimensional que engloba respostas cognitivas, emocionais, comportamentais, sensoriais e sociais manifestadas ao longo da totalidade da jornada de compra (Lemon & Verhoef, 2016). Utilizando a teoria da mente consciente como base teórica, a experiência do cliente define-se como as respostas mentais subjetivas, orientadas e multidimensionais de um cliente a uma interação com um parceiro de experiência, num *touchpoint* de uma etapa específica da jornada do cliente (Gahler et al., 2023). Esta conceptualização reconhece a valência subjetiva inerente às respostas mentais, permitindo aos clientes diferenciarem o grau de positividade das suas respostas através das múltiplas dimensões experienciais, enquanto as jornadas do cliente se posicionam como elemento fundamental de criação de valor, operacionalizando-se através de respostas multidimensionais a sequências coordenadas de *touchpoints* motivadas por objetivos específicos que integram pontos de contacto diretos e indiretos numa experiência coesa e holística (Y. T. J. Lin et al., 2023). Consequentemente, esta abordagem transcende a análise fragmentada de *touchpoints* individuais, priorizando a jornada completa como unidade de análise estratégica, onde a criação de valor se desenvolve continuamente ao longo do tempo através da coordenação sinérgica entre múltiplos pontos de contacto.

Esta conceptualização transcende a mera transação comercial, abrangendo todos os pontos de contacto onde o cliente interage com o negócio, produto ou serviço, numa perspetiva holística que reconhece que os consumidores procuram experiências significativas em detrimento de produtos isolados. O que as pessoas realmente desejam não são produtos, mas sim experiências que retiram durante a compra e após a compra (Lemon & Verhoef, 2016).

A experiência do cliente na jornada de compra encontra-se em constante evolução, com cada vez mais camadas de complexidade, como demonstra a progressão desde modelos iniciais de processo de decisão até à atual era de cocriação. A evolução histórica demonstra uma progressão paradigmática que se iniciou na criação de modelos de processo de decisão, evoluindo posteriormente para modelos de análise da reação do cliente destinados a perceber o nível de satisfação e lealdade (Olufunke Anne Alabi et al., 2024).

Subsequentemente, emergiu a preocupação com a qualidade do serviço, introduzindo a análise do ambiente e contexto onde a experiência de compra acontece. Esta fase foi fundamental para o desenvolvimento do mapeamento da jornada do cliente e a análise da qualidade do serviço em cada momento crítico da jornada (Lemon & Verhoef, 2016; Verhoef et al., 2009). A evolução prosseguiu com o marketing relacional, que se focou no processo de criação de confiança e compromisso com os clientes, analisando o custo emocional da mudança e resultando na investigação das emoções e percepções na experiência de compra. Esta transformação representa uma mudança paradigmática de foco puramente funcional para experiências emocionalmente envolventes, estabelecendo experiências hedónicas como drivers de diferenciação competitiva com benefícios comprovados incluindo maior propensão de compra, *word-of-mouth* (Warrington, 2002) positivo e sucesso empresarial sustentável (Ponsignon, 2023).

À medida que a proliferação de dados sobre os clientes se intensificava, as empresas começaram a perceber a necessidade de gerir a relação com o cliente de forma mais sofisticada (Papazafeiropoulou & Spanaki, 2016), compreendendo como determinados elementos específicos da experiência se relacionavam entre si e impactavam a rentabilidade do negócio. Com isto, surge a visão do marketing centrada no cliente, que transcendeu o domínio exclusivo do marketing para ser incorporada na cultura organizacional, onde toda a experiência de compra se centra em formas de melhorar a experiência antes, durante e após a compra (Lemon & Verhoef, 2016).

A transformação contemporânea posiciona o envolvimento do cliente como elemento central, estabelecendo processos de cocriação que se iniciam antes da definição de necessidades e se estendem através de comunidades digitais onde os consumidores partilham experiências, gerando influência social positiva proporcional à qualidade da experiência vivenciada. Verhoef et al., 2009 identificam que as experiências do cliente possuem um efeito cumulativo, onde as experiências passadas influenciam percepções atuais, criando um ciclo de expectativas crescentes que impulsiona a procura por uma personalização cada vez mais sofisticada. Quando um cliente tem uma boa experiência numa determinada empresa, tenderá a procurar essa mesma experiência no futuro, tornando-se mais exigente e procurando melhorar ainda mais a sua experiência. Esta associação positiva leva o cliente a procurar esse nível de experiência noutras empresas em jornadas de compra de diferentes produtos. Esta evolução de paradigma materializa-se através de quatro componentes interligados que estruturam a transformação experiencial. O *engagement emocional* (Boyatzis, 2002; Hollebeek, 2011) funciona como trigger para toda a jornada subsequente, criando conexão emocional desde o primeiro contacto e estabelecendo a fundação para capturar e manter a atenção do cliente. O processo simplificado visa reduzir fricção em todos os processos, mantendo o *momentum* emocional criado e facilitando a progressão na jornada. A individualização da experiência personaliza a experiência para necessidades específicas de cada cliente, funcionando como driver principal para decisões de compra e aumentando relevância e valor percebido. A sustentação ao longo do tempo mantém o nível de *engagement* consistente durante toda a jornada

através de pontos de contacto regulares e significativos, assegurando lealdade e satisfação de longo prazo (Hollebeek et al., 2016; Ponsignon, 2023).

2.2.2 FRAGMENTAÇÃO DAS MEDIDAS DE EXPERIÊNCIA DO CLIENTE

Existe atualmente uma fragmentação significativa das medidas utilizadas para avaliar a experiência do cliente, com escalas existentes específicas para contextos únicos, impossibilitando comparações de experiência entre diferentes interações e criando uma lacuna fundamental na compreensão *omnicanal* da experiência (Gahler et al., 2023).

A fragmentação representa um obstáculo crítico para as organizações contemporâneas que operam em ecossistemas com múltiplos canais, integrando *touchpoints online* e *offline* numa estratégia holística (Gahler et al., 2023). Esta desarticulação metodológica impede a otimização sistemática da experiência do cliente, particularmente relevante quando se considera que experiências positivas criam um ciclo de expectativas crescentes. (Verhoef et al., 2009) Evidenciam que experiências de qualidade superior elevam os padrões de exigência dos clientes, criando um *efeito de spillover* que transcende a empresa original e influencia expectativas em diferentes jornadas de compra, elevando progressivamente os padrões de experiência no mercado.

Esta fragmentação exige uma abordagem integrada que unifique a recolha e análise de dados através de todos os touchpoints e interações, criando uma visão holística da experiência do cliente. A recolha estratégica de dados constitui um pilar fundamental da experiência do cliente contemporânea, permitindo que algoritmos preditivos e ferramentas de IA antecipem necessidades e personalizem cada momento da jornada de compra (Rusnaini et al., 2024). Para maximizar esta capacidade preditiva, os dados devem ser consistentemente recolhidos e integrados através de múltiplas dimensões, nomeadamente diferentes parceiros de experiência (marca, funcionários, outros clientes), *touchpoints omnicanal* (*offline* e *online*), todas as etapas da jornada (pré-compra, compra, pós-compra) e diversos setores industriais (Gahler et al., 2023). Esta abordagem holística de gestão de dados transforma informação fragmentada em *insights* acionáveis que elevam o acompanhamento personalizado do cliente (Roy et al., 2025).

Esta evolução conceptual revela tensões fundamentais que caracterizam o estado atual da investigação da experiência do cliente. Emerge um paradoxo entre a sofisticação teórica alcançada e a fragmentação das ferramentas de medição, onde a progressão desde modelos básicos (L. Berry, 2002; Fournier, 1998) de decisão até à era da cocriação contrasta com a impossibilidade de comparar experiências entre diferentes contextos. Simultaneamente, observa-se uma tensão entre a necessidade de personalização crescente, impulsionada pelo ciclo de expectativas crescentes, e a exigência de standardização metodológica que permita análises comparativas sistemáticas. A complexidade multidimensional da experiência do cliente, embora teoricamente rica, cria desafios operacionais para a gestão prática, particularmente quando se considera que o *efeito spillover* (Wagner, 1891) transcende o

controle direto por parte da organização. Perante estas tensões conceituais, a inteligência artificial emerge como elemento mediador crucial, oferecendo capacidades únicas para conciliar personalização com standardização através de algoritmos que processam múltiplas dimensões simultaneamente (Gandhudi et al., 2025), mantendo comparabilidade metodológica enquanto automatizam análises complexas em interfaces operacionalmente simples (V. Kumar et al., 2019; Rusnaini et al., 2024). Esta convergência tecnológica posiciona a IA não apenas como ferramenta de otimização, mas como uma solução integradora que pode superar a atual fragmentação das medidas da experiência do cliente (Ledro et al., 2025).

2.2.3. PERSONALIZAÇÃO: DIMENSÕES MULTIDIMENSIONAIS E PARADOXOS

A personalização emerge como paradigma central na evolução do serviço ao cliente, representando uma mudança fundamental do marketing de massa para abordagens centradas no indivíduo (Chandra et al., 2022). A personalização é definida como um processo tecnológico automatizado que adapta conteúdo baseado em dados do cliente (Lambillotte et al., 2022). Enquanto na perspectiva de marketing, a personalização é conceptualizada como um processo estratégico para criar interações individualizadas e relevantes que melhoram significativamente a experiência do cliente (Chandra et al., 2022). Esta dualidade conceptual revela a natureza híbrida da personalização contemporânea, que combina capacidades tecnológicas avançadas com objetivos estratégicos de marketing, criando um ecossistema onde a automação tecnológica serve como meio para alcançar relevância experiencial. A convergência destas duas perspectivas materializa-se na capacidade de processar dados em tempo real para gerar interações que não apenas respondem às necessidades imediatas do cliente, mas antecipam preferências futuras (Gandhudi et al., 2025), estabelecendo um ciclo contínuo de aprendizagem e adaptação que eleva progressivamente a qualidade da experiência personalizada (Dong & McIntyre, 2014).

As descobertas recentes indicam que a personalização dos *touchpoints* afeta todas as cinco dimensões de resposta da experiência do cliente através de um espectro abrangente de cinco tipos de resposta distintos. As respostas *cognitivas* manifestam-se tanto positivamente através da utilidade percebida, relevância da oferta e qualidade de decisão, como negativamente mediante preocupações com a privacidade, ceticismo e percepção de manipulação. Paralelamente, as respostas *emocionais* oscilam entre sentimentos positivos de cuidado, prazer e diversão, e reações negativas como vulnerabilidade percebida, medo de exploração e irritação (Quach et al., 2022). As respostas *comportamentais* evidenciam-se no aumento de cliques e abertura de emails versus rejeição de anúncios e reatância, enquanto as respostas *sensoriais* concentram-se predominantemente em maior fixação visual e atenção. As respostas *sociais* incluem presença social e familiaridade, contrastando com embaraço em contextos públicos, evidenciando o paradoxo da personalização onde benefícios coexistem sistematicamente com preocupações (Weidig et al., 2024).

Esta complexidade multidimensional varia conforme a natureza dos *touchpoints*: os *touchpoints digitais* focam nas respostas cognitivas e comportamentais com personalização em tempo real, mas

geram preocupações de privacidade; os *touchpoints humanos* privilegiam respostas emocionais e sociais com alta flexibilidade, mas limitações de automatização; e os *touchpoints físicos* enfatizam as respostas cognitivas e sociais com familiaridade, mas estáticos. Cada tipo de resposta, manifesta-se de diferentes formas ao longo da jornada do cliente, desde facilitar decisões pré-compra até fomentar proximidade pós-compra. Para maximizar esta capacidade preditiva e transformacional, os dados devem ser consistentemente recolhidos através de métodos explícitos e implícitos. Integrados através de múltiplas dimensões abrangendo diferentes parceiros de experiência e *touchpoints* ao longo de todas as etapas da jornada, permitindo uma mudança paradigmática de foco puramente funcional para experiências emocionalmente envolventes que estabelecem experiências hedónicas como drivers de diferenciação competitiva com benefícios comprovados incluindo maior propensão de compra, *word-of-mouth* positivo e crescimento comercial sustentável (Gahler et al., 2023; Ponsignon, 2023).

A revisão da literatura identifica seis temas principais na investigação sobre personalização, nomeadamente *recomendação personalizada* através de sistemas baseados em IA e ML, *relacionamento personalizado* via a gestão do relacionamento do cliente individualizado, o *paradoxo personalização-privacidade* que equilibra benefícios com preocupações de confidencialidade, *publicidade personalizada* com segmentação avançada usando *big data*, *conceitos e discursos teóricos de personalização no marketing*, e *insights comportamentais* para prever necessidades e preferências (Chandra et al., 2022). Esta taxonomia revela simultaneamente a maturidade e a fragmentação do campo de investigação em personalização. Enquanto os domínios de recomendação e publicidade personalizada beneficiam de avanços tecnológicos consolidados através de algoritmos *de machine learning* cada vez mais sofisticados, o paradoxo personalização-privacidade (Hongladarom, 2023) permanece como tensão central não resolvida, evidenciando a necessidade premente de *frameworks* éticos mais robustos que conciliem eficácia comercial com proteção dos dados pessoais. A convergência emergente entre *insights* comportamentais e relacionamento personalizado sugere oportunidades significativas para abordagens mais holísticas que transcendam a mera segmentação demográfica, contudo, a literatura ainda revela lacunas substanciais na ausência de modelos integrativos que unifiquem estas dimensões numa perspetiva sistémica da personalização da experiência do cliente (Roy et al., 2025).

A resolução deste paradoxo e a maximização dos benefícios da personalização exigem uma integração estratégica sofisticada da IA, que emerge não como tecnologia separada, mas como *enabler* fundamental que permite transcender as limitações tradicionais da personalização (Dong & McIntyre, 2014). A IA transforma-se assim no mecanismo que viabiliza a personalização em escala massiva, resolvendo simultaneamente os desafios de complexidade multidimensional, processamento de dados em tempo real e otimização contínua da experiência do cliente, enquanto mitiga preocupações de privacidade através de algoritmos mais transparentes e controláveis (Hermann & Puntoni, 2024).

2.2.4. A IA COMO *ENABLER* DA PERSONALIZAÇÃO

Sodiq Odetunde Babatunde et al., 2024 estabelecem que a personalização baseada em IA aumenta significativamente o *engagement*, satisfação, retenção e vendas através de três teorias fundamentais, nomeadamente o *Modelo de Probabilidade de Elaboração* (MPE) (Petty & Cacioppo, 1986), que demonstra como mensagens personalizadas geram maior elaboração cognitiva, a *Teoria dos Usos e Gratificações* (TUG) (E. Katz et al., 1973), onde IA identifica necessidades específicas para entregar conteúdo relevante, e a *Teoria Cognitiva Social* (TCS) (McLeod, 2025), que utiliza prova social e auto-eficácia para reforçar conexões com a marca. (V. Kumar et al., 2019) complementam esta perspetiva ao identificar que a IA revoluciona o marketing através da personalização e curadoria de conteúdo, distinguindo personalização (empresa decide baseada em dados) de customização (cliente especifica proativamente), enquanto potencializa ambas através de análise de dados em larga escala e aprendizagem contínua, criando assim um ecossistema integrado onde as capacidades tecnológicas convergem com os fundamentos teóricos do comportamento do consumidor para maximizar a eficácia das estratégias de *engagement* e conversão (Xing et al., 2023).

Esta convergência teórica adquire impacto estratégico particular nas MSPs, uma vez que a conectividade e o efeito rede constituem pilares fundamentais de criação de valor (Mlčúchová, 2022; Pousttchi & Gleiss, 2019; Sanchez-Cartas & León, 2021). As três teorias amplificam sistematicamente estes pilares, em que o MPE intensifica a qualidade das conexões, a TUG aumenta a adoção e retenção ao entregar conteúdo personalizado aos utilizadores, e a TCS acelera o crescimento da rede através de recomendações inteligentes baseadas no comportamento da comunidade.

A integração da IA na personalização representa uma transformação paradigmática que transcende a mera implementação tecnológica, constituindo uma reconfiguração fundamental dos modelos de relacionamento entre marcas e consumidores no ecossistema digital contemporâneo (Hongladarom, 2023). Esta mudança manifesta-se através da capacidade dos algoritmos de IA para processar e analisar *datasets* vastos e complexos, extraindo *insights* comportamentais ocultos que as abordagens tradicionais de segmentação e análise não conseguem capturar devido às suas limitações metodológicas inerentes (Babatunde et al., 2024). A personalização com IA emerge assim como catalisador de conexões mais profundas e significativas com os clientes, permitindo a criação de algoritmos preditivos sofisticados que não respondem apenas às necessidades expressas, mas antecipam preferências latentes através da análise contínua de padrões comportamentais (Chen & Prentice, 2025).

Esta capacidade transformacional exige uma abordagem holística que integre todos os *touchpoints* da experiência do cliente numa perspetiva sistémica, aplicando-se longitudinalmente ao longo de toda a jornada do cliente - desde a fase pré-compra de descoberta e consideração, passando pelo momento transacional, até às interações pós-compra de retenção e *advocacy* - através de um *framework* conceptual estruturado que assegure coerência experiencial e maximize o potencial de criação de relacionamentos duradouros e mutuamente benéficos entre marcas e consumidores. Esta abordagem holística

operacionaliza-se através de três dimensões fundamentais que estruturam a implementação da IA na experiência do cliente, nomeadamente AI Experience (Yang, 2017) (como os clientes experienciam e percebem as interações com IA), AI Functions (Russell et al., 1995) (diferentes capacidades e funcionalidades da IA aplicadas ao serviço) e AI Services (Kitsara, 2022) (tipos específicos de serviços habilitados por IA e seu impacto).

Com a personalização, a experiência do cliente, no *cross-selling*, transcende a tradicional venda de produtos adicionais para se transformar numa estratégia disruptiva de criação de valor mútuo (Ghoshal et al., 2021), onde a IA redefine fundamentalmente a relação comercial através da antecipação inteligente de necessidades não expressas pelo cliente (Chen & Prentice, 2025; Hermann & Puntoni, 2024). Esta abordagem revolucionária materializa-se quando, após a seleção de um produto focal, sistemas de ML não apenas oferecem produtos complementares baseados em preferências históricas, mas criam ecossistemas de valor personalizados que geram benefícios exponenciais tanto para empresas (aumentos de receita até 20%, como demonstrado por implementações baseadas em conselhos McKinsey) quanto para clientes (maximização da utilidade através de soluções integradas que não teriam descoberto autonomamente) (Ghoshal et al., 2021).

A verdadeira disrupção emerge da capacidade da IA de transformar cada transação numa oportunidade de cocriação de valor, onde a personalização avançada, análise preditiva comportamental e otimização dinâmica em tempo real convergem para criar experiências comerciais que antecipam desejos latentes, redefinindo o *cross-selling* de tática de vendas para arquitetura de relacionamento inteligente que maximiza simultaneamente a eficácia comercial e a satisfação experiencial do cliente (Chen & Prentice, 2025; Lemon & Verhoef, 2016).

2.2.5. SÍNTESE CRÍTICA E CONVERGÊNCIAS PARADIGMÁTICAS

A análise da personalização da experiência do cliente através da IA revela um cenário complexo onde tensões conceituais fundamentais coexistem com sinergias transformacionais, evidenciando simultaneamente choques paradigmáticos e convergências estratégicas que moldam o futuro do marketing relacional.

As principais tensões emergem entre a sofisticação teórica alcançada na conceptualização multidimensional da experiência do cliente (Lemon & Verhoef, 2016) e a persistente fragmentação das ferramentas de medição (Gahler et al., 2023). Este paradoxo operacional onde a capacidade de personalização avançada contrasta com a impossibilidade de comparar experiências entre diferentes contextos, enquanto o ciclo de expectativas crescentes (Verhoef et al., 2009) impulsiona uma procura por personalização cada vez mais sofisticada que colide com preocupações éticas crescentes sobre privacidade (Quach et al., 2022) e manipulação algorítmica evidenciadas nas respostas cognitivas negativas dos clientes (Weidig et al., 2024). Contudo, estas tensões intensificam-se precisamente devido à convergência entre as capacidades tecnológicas da IA e os fundamentos teóricos do comportamento

do consumidor (Babatunde et al., 2024), onde o Modelo de Probabilidade de Elaboração, a Teoria dos Usos e Gratificações, e a Teoria Cognitiva Social se alinham perfeitamente com algoritmos de ML para criar um ecossistema integrado que transforma cada transação numa oportunidade de cocriação de valor (Chen & Prentice, 2025; Ghoshal et al., 2021), materializando-se através da evolução do *cross-selling* de tática de vendas para arquitetura de relacionamento inteligente.

Apesar das tensões inerentes, todas as perspectivas convergem no reconhecimento da personalização em conjunto com a IA como benefícios fundamentais para o futuro do marketing (V. Kumar et al., 2019; Mukhopadhyay et al., 2024). Contudo, a evolução futura exige maturação significativa em múltiplas dimensões críticas, nomeadamente o desenvolvimento de competências analíticas avançadas que transcendam a mera implementação tecnológica (Anshari et al., 2019), a curadoria de conteúdos que equilibre relevância com ética, e o alinhamento organizacional transversal que supere hierarquias tradicionais (Kumar et al., 2019). A resolução das questões éticas e de privacidade através de *frameworks* robustos de transparência e controlo do utilizador (Babatunde et al., 2024), e o refinamento dos três estágios lógicos da personalização - recolha de dados, análise e apresentação de *insights* (Wai Wong et al., 2004) - para suportar técnicas de venda *one-to-one* em escala massiva como evolução natural do marketing relacional, com benefícios comprovados que incluem melhoria significativa na (i) compreensão e previsão de comportamento do consumidor, (ii) estratégias de *targeting* mais eficazes, (iii) aumento do *engagement* e fidelização, (iv) mensagens de marketing mais relevantes e oportunas (Alijoyo et al., 2025; Xing et al., 2023). Criando assim relacionamentos autênticos através de tecnologia inteligente que maximiza simultaneamente a eficácia comercial, a satisfação e o *engagement* (Ledro et al., 2025; Sheth et al., 2023; Xing et al., 2023), posicionando a convergência IA-personalização não apenas como tendência tecnológica, mas como mudança fundamental na forma como as marcas se conectam com os consumidores no contexto digital em evolução (Chen & Prentice, 2025).

As aplicações práticas da personalização baseada em IA demonstram impactos mensuráveis significativos. Boudet et al., (2019) documentam aumentos de receita entre 5-15% e melhorias de eficiência de marketing de 10-30%, principalmente através de recomendações de produtos e comunicações acionadas. A segmentação avançada transcende demografia básica, utilizando múltiplos pontos de dados para identificar padrões comportamentais complexos e criar mensagens altamente direcionadas (Alijoyo et al., 2025). No setor segurador, esta convergência entre personalização e IA encontra a sua expressão mais disruptiva através das MSPs, que emergem como arquiteturas tecnológicas capazes de operacionalizar, em grande escala, os princípios de marketing relacional e personalização da experiência do cliente anteriormente analisados (Mukhopadhyay et al., 2024). Estas plataformas representam não apenas uma inovação tecnológica, mas uma reconfiguração fundamental dos modelos de intermediação no setor segurador, criando ecossistemas digitais onde a personalização se torna sistemática e os modelos de ML transformam cada interação numa oportunidade de cocriação de valor (Anshari et al., 2019).

2.3. MULTI-SIDED PLATFORMS NO SETOR SEGURADOR

2.3.1. DEFINIÇÃO E ARQUITETURA DAS MSPS

As multi-sided platforms (MSPs) constituem modelos de negócio fundamentalmente disruptivos que facilitam interações diretas entre dois ou mais grupos distintos de utilizadores, criando valor através da conectividade e dos efeitos de rede (Mlčúchová, 2022; Pousttchi & Gleiss, 2019; Sanchez-Cartas & León, 2021). No contexto do setor segurador, estas plataformas emergem como intermediários tecnológicos revolucionários que transformam radicalmente a distribuição e comercialização tradicional de seguros, introduzindo paradigmas de criação de valor anteriormente inexistentes (Pousttchi & Gleiss, 2019).

As MSPs tendem a prosperar quando existe assimetria de informação (Chan et al., 2022) e o setor dos seguros, nesse sentido, tem características únicas. É um produto percecionado pelos clientes como de alto risco, porque não existe total certeza por parte do cliente de poder acionar a cobertura do seguro que contratou, o que faz com que a confiança seja um elemento-chave na relação com os clientes (Méndez-Aparicio et al., 2020). Os clientes possuem dois perfis comportamentais distintos, em que no momento da contratação o cliente é racional e avalia a informação e segurança que o produto possui, e é emocional na utilização do serviço, uma vez que quando acontece um sinistro ou perdas de bens ou pessoas, a pessoa encontra-se num estado muito emotivo (Méndez-Aparicio et al., 2020).

A arquitetura fundamental das MSPs assenta em três pilares estruturais que as distinguem categoricamente dos modelos lineares tradicionais (Alexander et al., 2023; Allweins et al., 2021; Suuronen et al., 2024):

1. A facilitação de interações diretas onde a plataforma atua como orquestrador tecnológico sem participar diretamente nas transações;
2. A afiliação estratégica que representa a decisão consciente dos utilizadores de se integrarem no ecossistema da plataforma;
3. A criação de valor multidimensional através da redução de custos de transação, melhoria da eficiência de *matching*, e disponibilização de informação que facilita a tomada de decisões informadas.

2.3.2. TRANSFORMAÇÃO DA PERCEÇÃO DA CRIAÇÃO DE VALOR

A transformação da perceção de criação de valor representa outro aspeto fundamental da importância das MSPs. Tradicionalmente, o valor no setor segurador era criado principalmente através da gestão de ativos, subscrição de riscos e gestão de sinistros (McFall et al., 2020). As MSPs introduzem novas dimensões de valor, focando na experiência do cliente, conveniência, transparência e personalização (Pousttchi & Gleiss, 2019).

Os elementos de criação de valor manifestam-se através de quatro dimensões críticas, nomeadamente: (i) ambiente confiável que mitiga assimetrias informacionais endémicas no setor

segurador, (ii) expansões baseadas em dados que permitem personalização sofisticada impossível nos modelos tradicionais, (iii) serviços personalizados que respondem dinamicamente às necessidades específicas de cada utilizador, e (iv) mecanismos de envolvimento que sustentam a participação ativa ao longo de toda a jornada do cliente (Suuronen et al., 2024). Estas dimensões fundamentam uma mudança paradigmática na forma como o valor é criado e entregue aos clientes. Esta transformação não é meramente incremental - representa uma revolução na conceção de valor que altera as expectativas dos consumidores e força o setor tradicional a repensar integralmente as suas propostas (Pousttchi & Gleiss, 2019).

A superioridade das MSPs na criação de valor manifesta-se através de múltiplas dimensões, nomeadamente redução dramática de custos de pesquisa que permite aos utilizadores encontrar facilmente produtos adequados às suas necessidades (Micallef et al., 2023), melhoria exponencial da qualidade de *matching* através de algoritmos de ML que conectam utilizadores com base em preferências específicas (Alexander et al., 2023; Allweins et al., 2021), disponibilização de informação transparente e comparável que facilita decisões verdadeiramente informadas (Allweins et al., 2021), e criação de confiança através de sistemas de *rating e reviews* que reduzem substancialmente a incerteza nas transações (Chan et al., 2022).

2.3.3. IMPORTÂNCIA DAS MSPS NO SETOR SEGURADOR

Estabelecidos os fundamentos arquiteturais e os mecanismos de criação de valor das MSPs, torna-se imperativo analisar como estas características estruturais se traduzem em vantagens competitivas específicas e transformações setoriais mensuráveis no contexto do mercado segurador português e internacional. A relevância estratégica das MSPs transcende a sua elegância conceptual, manifestando-se através de impactos disruptivos concretos que redefinem a dinâmica competitiva, as expectativas dos consumidores e os modelos de rentabilidade no setor segurador (Parker et al., 2016).

O setor segurador enfrenta desafios estruturais significativos que tornam as MSPs particularmente relevantes como veículos de inovação e transformação. A rentabilidade do setor tem sido pressionada por fatores como baixas taxas de juro, aumento da sinistralidade em alguns ramos, e intensificação da concorrência. Adicionalmente, existe uma carência de estudos académicos sobre a aplicação de MSPs especificamente no setor segurador, criando uma oportunidade de investigação significativa (McFall et al., 2020; Pousttchi & Gleiss, 2019).

As MSPs operam com uma visão *customer centric* (Hanelt et al., 2021) que contrasta radicalmente com a abordagem *product-centric* (Christensen et al., 2016) das seguradoras. Esta diferença fundamental manifesta-se na forma como as MSPs veem o cliente como o centro do ecossistema, não como um recetor passivo de produtos pré-definidos (Méndez-Aparicio et al., 2020). A transformação digital através das MSPs é impulsionada pela relação dinâmica entre expectativas crescentes dos clientes (Verhoef et al., 2009) e a qualidade percebida dos serviços (Lemon & Verhoef, 2016), onde as MSPs

procuram sistematicamente eliminar o gap entre expectativas elevadas e experiências tradicionais deficitárias. Esta orientação *customer centric* traduz-se em estruturas organizacionais permeáveis e ágeis, com uma gestão focada que é suportada por tecnologia, e possui processos automatizados baseados em dados. O *customer-centricity* emerge como pilar fundamental do ecossistema digital, onde a experiência do cliente é sistematicamente otimizada para minimizar o esforço do cliente e maximizar os níveis de satisfação através de cada *touchpoint* da jornada (Sheth et al., 2023).

A vantagem competitiva mais significativa das MSP reside na sua capacidade superior de acesso, recolha e utilização de dados que transcende as fronteiras organizacionais tradicionais. Enquanto as seguradoras convencionais possuem informação detalhada apenas sobre os seus próprios clientes e produtos, as MSPs têm acesso a dados transversais sobre comportamentos de compra, preferências e necessidades de seguros de uma base alargada de utilizadores, permitindo uma compreensão holística do mercado impossível de alcançar através de modelos tradicionais (Parker et al., 2016). Esta superioridade informacional traduz-se diretamente num impacto transformador na jornada de compra de seguros, através da introdução de maior transparência via comparações diretas entre produtos e preços, simplificação do processo de cotação e contratação, personalização de recomendações baseada em perfis individuais, e melhoria da experiência global do cliente através de interfaces intuitivas e processos digitalizados. Consequentemente, esta transformação resulta em maior satisfação do cliente e potencial aumento da penetração de seguros no mercado, criando um ciclo virtuoso de crescimento e inovação no setor (Pousttchi & Gleiss, 2019).

A transformação digital trouxe alterações profundas ao comportamento do consumidor. O consumidor digital está mais conectado, procura ativamente por informação antes de comprar, e exige personalização sofisticada das experiências ao longo da jornada de compra (Verhoef et al., 2021). Para responder a estas expectativas crescentes, a evolução da IA levou ao desenvolvimento de algoritmos sofisticados que abordam problemas complexos do mundo real de forma mais eficaz. Estes avanços têm sido particularmente impactantes no setor financeiro e segurador, onde as técnicas de ML permitiram um processamento de dados mais eficiente, análise perspicaz e, crucialmente, a personalização em massa das ofertas aos consumidores (Alexander et al., 2023).

2.3.4. SÍNTESE CRÍTICA E CONVERGÊNCIAS PARADIGMÁTICAS

A análise das MSPs no setor segurador revela uma convergência paradigmática extraordinária entre três correntes teóricas fundamentais que se articulam numa síntese emergente de elevada sofisticação analítica. A teoria das plataformas digitais (Mlčúchová, 2022; Pousttchi & Gleiss, 2019; Teece et al., 2022), a teoria da cocriação de valor (Hanelt et al., 2021; Payne & Frow, 2006) e a teoria dos efeitos de rede (Pousttchi & Gleiss, 2019; Sanchez-Cartas & León, 2021) não constituem meramente *frameworks* complementares, e sim elementos constitutivos de um meta-paradigma explicativo que transcende as limitações das abordagens unidimensionais tradicionais. Esta convergência manifesta-se através de uma arquitetura conceptual tripartida (Alexander et al., 2023; Allweins et al., 2021; Suuronen et al., 2024)

que integra sistematicamente as dimensões tecnológica, comportamental e económica, estabelecendo *loops de feedback* positivos onde a participação crescente gera melhores dados, que por sua vez possibilitam uma personalização superior (Pousttchi & Gleiss, 2019), incrementando o valor percebido e atraindo mais utilizadores num ciclo virtuoso autorreforçado. A especificidade do setor segurador, caracterizada por assimetrias informacionais endémicas (Chan et al., 2022; Méndez-Aparicio et al., 2020) e dualidade comportamental dos consumidores (Méndez-Aparicio et al., 2020), confere particular relevância a esta síntese teórica, uma vez que as MSPs emergem como orquestradores tecnológicos capazes de resolver tensões estruturais através da combinação sinérgica de eficiência algorítmica e sensibilidade às necessidades humanas complexas.

A contextualização setorial específica do setor segurador evidencia como as características únicas deste mercado - nomeadamente a centralidade da confiança (Méndez-Aparicio et al., 2020), a perceção de alto risco (Méndez-Aparicio et al., 2020), a dualidade comportamental (Méndez-Aparicio et al., 2020) e a complexidade da assimetria informacional (Chan et al., 2022) - influenciam decisivamente as dinâmicas competitivas e os modelos de criação de valor das MSPs (Pousttchi & Gleiss, 2019).

A evidência empírica converge numa realidade incontornável, as MSPs representam uma transformação paradigmática que transcende meras tendências tecnológicas. Esta mudança manifesta-se através da reconfiguração fundamental da arquitetura competitiva do setor segurador, onde as MSPs emergem como intermediários de conhecimento que redefinem irreversivelmente as relações de poder entre *stakeholders* (Allweins et al., 2021). A combinação da superioridade estrutural na criação de valor, visão *customer centric*, supremacia no acesso e utilização de dados, algoritmos preditivos, cocriação de valor, e capacidades de *cross-selling*, cria vantagens competitivas que tornam a transformação do setor inevitável (Patterson et al., 2014). Contudo, esta transformação não é isenta de riscos sistémicos. A concentração de poder de mercado, vulnerabilidades de cibersegurança, questões críticas de privacidade de dados, e potenciais falhas de mercado decorrentes de assimetrias informacionais persistentes constituem desafios estruturais significativos (Quach et al., 2022). Adicionalmente, os *frameworks* regulamentares específicos do setor segurador, tanto nacionais quanto internacionais, podem limitar ou condicionar a velocidade e amplitude desta transformação.

A capacidade transformadora das MSPs materializa-se de forma particularmente evidente através da otimização de estratégias de *cross-selling* (Patterson et al., 2014), onde a convergência entre dados comportamentais privilegiados, algoritmos preditivos e personalização cria oportunidades comerciais anteriormente inalcançáveis pelos modelos tradicionais de distribuição seguradora. Esta sinergia tecnológico permite às MSPs transcender as limitações históricas do *cross-selling*, caracterizado por baixas taxas de conversão e abordagens generalizadas, transformando-o numa arquitetura de relacionamento inteligente que antecipa necessidades latentes e maximiza simultaneamente valor para clientes e organizações (Kwak et al., 2019).

2.4. CROSS-SELLING E ALGORITMOS PREDITIVOS NO SETOR SEGURADOR

2.4.1. CROSS-SELLING: FUNDAMENTOS CONCEPTUAIS E ESTRATÉGICOS

O *cross-selling* constitui uma arquitetura comercial que consiste na comercialização de produtos e serviços adicionais, invariavelmente complementares, subsequentemente à aquisição de um produto focal (Ghoshal et al., 2021). No contexto do ecossistema segurador, esta estratégia manifesta-se através da promoção de produtos de seguro adicionais para segurados preexistentes (Akhavan & Hassannayebi, 2024). Quando operacionalizado de forma adequada, o *cross-selling* permite comercializar produtos que correspondem genuinamente às necessidades latentes dos clientes, gerando valor mútuo mediante benefícios bidirecionais significativos (Tolga Akçura & Srinivasan, 2005).

Para as organizações, esta arquitetura estratégica proporciona a maximização da receita por cliente sem custos adicionais de aquisição, otimização da rentabilidade através de diversificação dos riscos, incremento da retenção decorrente da dependência de múltiplos produtos, e aprofundamento do conhecimento comportamental do cliente. Para os clientes, os benefícios incluem conveniência na gestão centralizada, vantagens económicas via descontos por fidelidade, e melhor coordenação entre coberturas (Li et al., 2011; Tolga Akçura & Srinivasan, 2005). Por conseguinte, o *cross-selling* estrategicamente orquestrado contribui para a redução do *churn*, aumento da lealdade, e maximização do *customer lifetime value*, estabelecendo um ciclo virtuoso que consolida a posição competitiva no mercado.

Uma estratégia de *cross-selling* inteligente visa maximizar o CLV, sustentando-se no pressuposto de que clientes satisfeitos apresentam maior propensão para aquisições adicionais, em que (Verhoef et al., 2009) identificaram correlações positivas estatisticamente significativas entre os níveis de satisfação inicial e a recetividade subsequente a ofertas complementares. Do ponto de vista económico, os custos de aquisição marginais para clientes existentes revelam-se substancialmente inferiores aos custos associados à captação de novos clientes (Kumar & Reinartz, 2016) quantificada por (Kumar & Reinartz, 2016), que demonstraram empiricamente que o *cross-selling* pode reduzir os custos de aquisição até 70%, enquanto simultaneamente incrementa a rentabilidade por cliente numa amplitude de 15-25%. Paralelamente, emerge uma dimensão educacional inerente à apresentação de produtos complementares (Li et al., 2011), que contribui para o desenvolvimento progressivo da literacia financeira dos clientes através da exposição sistemática a conhecimento especializado de forma positiva e recorrente (Bhatia et al., 2021; Fisch & Seligman, 2022), posicionando assim o *cross-selling* não apenas como uma tática comercial, mas como uma estratégia integrada de criação de valor mútuo entre empresa e cliente.

2.4.2. TENSÕES TEÓRICAS E TRANSFORMAÇÃO PARADIGMÁTICA

Apesar dos benefícios potenciais, o *cross-selling* enfrenta desafios significativos que limitam a sua eficácia. O principal desafio reside na taxa de sucesso extremamente reduzida destas campanhas, tornando frequentemente estas iniciativas não rentáveis (Li et al., 2011). Esta baixa taxa de conversão

resulta da dificuldade em identificar o momento ótimo e o produto apropriado para cada cliente específico.

Adicionalmente, nem todos os clientes que adquirem múltiplos produtos se revelam lucrativos para a empresa. (Shah et al., 2012) identificaram que, apesar de comprarem vários produtos, alguns clientes exibem comportamentos adversos no pós-compra, incluindo reversões de receita excessivas (devoluções, cancelamentos, *chargebacks*) e pedidos de serviço desproporcionais. Paralelamente, Li et al., (2011) identificam que o *cross-selling* bem-sucedido contribui para a redução do *churn*, aumento da lealdade dos clientes e maximização do *customer lifetime value*, particularmente quando se adota uma perspectiva de longo prazo que reconhece o valor dos efeitos educacionais que transcendem a venda imediata. Tolga Akçura & Srinivasan, (2005) complementam esta visão, enfatizando que o *cross-selling* eficaz requer intimidade com o cliente e estratégias baseadas em compromisso que estabelecem confiança como (Hu et al., 2020) valiosa no relacionamento cliente-empresa (L. Berry, 2002). Esta abordagem sugere que o sucesso do *cross-selling* depende não apenas da identificação de oportunidades de venda, mas da compreensão profunda dos padrões comportamentais dos clientes e da implementação de estratégias que equilibrem a rentabilidade imediata com o valor do relacionamento de longo prazo (Silva-Atencio, 2025).

A qualidade e disponibilidade de dados constituem outro desafio crítico (Stourm et al., 2020). Existem lacunas significativas nos *datasets* da indústria seguradora, onde a falta de métodos padronizados de recolha de dados e disparidades nos formatos dificultam o desenvolvimento de modelos de *machine learning* precisos por parte das seguradoras (Akhavan & Hassannayebi, 2024).

2.4.3. INFLUÊNCIA DA PERSONALIZAÇÃO NO *CROSS-SELLING*

A personalização emerge como fator crítico para o sucesso das campanhas de *cross-selling*, requerendo uma abordagem *customer-centric* que reconhece a evolução dos clientes através de diferentes estados de maturidade financeira (Li et al., 2011).

A convergência entre os *insights* de Li et al., (2011) e Tolga Akçura & Srinivasan, (2005) revelam uma evolução estratégica fundamental no *cross-selling* contemporâneo, onde a competição pela atenção do cliente transforma-se numa oportunidade de construção de relacionamentos duradouros. A transição de uma abordagem *campaign-oriented* (Strömbäck, 2007) para *customer-centric*, encontra suporte nas estratégias baseadas em compromisso (Bhattacharya & Sen, 2003; Peppers & Rogers, 2016).

Enquanto as empresas tradicionalmente competiam pela atenção dos clientes através de campanhas pontuais e abordagens transacionais, o novo paradigma estratégico exige o desenvolvimento de intimidade genuína com o cliente como alicerce fundamental para um *cross-selling* inteligente. A orientação de tratar a confiança como moeda valiosa (Tolga Akçura & Srinivasan, 2005) adquire relevância crítica no contexto regulatório contemporâneo, marcado por rigorosas políticas de privacidade de dados (Quach et al., 2022) e conformidade com o *RGPD* (Papazafeiropoulou & Spanaki,

2016). Esta transformação *customer-centric* transcende a mera recolha de informação, exigindo que as empresas demonstrem valor tangível e transparente em contrapartida pela informação voluntariamente partilhada pelos clientes (Linoff & Berry, 2011), criando assim um ciclo virtuoso onde a confiança mútua alimenta uma personalização cada vez mais refinada (Li et al., 2011), que por sua vez fortalece o relacionamento e gera oportunidades de *cross-selling* autênticas e mutuamente benéficas.

A apresentação de produtos de forma personalizada deve evoluir dinamicamente com o cliente, integrando múltiplos canais de comunicação estrategicamente selecionados para cada momento específico e contexto relevante ao longo do ciclo de vida do relacionamento. (Li et al., 2011) demonstraram que o efeito educacional resultante deste acompanhamento personalizado e contínuo domina largamente outros efeitos mensuráveis, construindo um valioso *goodwill* e estabelecendo relacionamentos profundos que transcendem significativamente a transação imediata, criando uma base sólida de criação de valor de longo prazo (Patterson et al., 2014). Que vai beneficiar tanto o cliente, através de soluções mais adequadas às suas necessidades, quanto a empresa através de maior lealdade e *customer lifetime value*.

2.4.4. MACHINE LEARNING: ALGORITMOS PREDITIVOS SUPERVISIONADOS

Os algoritmos preditivos de classificação supervisionada têm demonstrado eficácia significativa na predição de oportunidades de *cross-selling* no setor segurador. Estas técnicas representam uma evolução das abordagens estatísticas tradicionais, oferecendo maior flexibilidade e automatização na construção de modelos (Hu et al., 2020).

A literatura identifica que os algoritmos preditivos de classificação supervisionada são os mais relevantes para aplicações de *cross-selling* no setor segurador (Mahadevkar et al., 2022). Caracterizam-se pelo treino com dados rotulados, esta categoria permite aos algoritmos de ML aprender a partir de dados de treino e desenvolver predições para novas observações (Hermann & Puntoni, 2024). No contexto do *cross-selling*, os algoritmos aprendem a partir de dados históricos de clientes que adquiriram ou não produtos específicos, desenvolvendo capacidade preditiva para novos casos (Obi, 2023; Szabó et al., 2024).

Hu et al., 2020 estabelecem três dimensões fundamentais que explicam a crescente adoção de algoritmos preditivos de classificação supervisionada no setor segurador, configurando elementos estruturantes da transformação digital contemporânea desta indústria. **(i) Processamento de Grandes Volumes de Dados** (Fayyad et al., 1996), em que a flexibilidade algorítmica inerente aos métodos de ML permite transcender as limitações das técnicas de modelação paramétricas convencionais, que se revelam constrangidas na sua capacidade de capturar relações complexas entrada-saída de natureza não-linear. Esta capacidade adaptativa assume relevância crítica no contexto segurador, onde os padrões comportamentais dos clientes manifestam características intrinsecamente não-lineares e multidimensionais, exigindo abordagens metodológicas capazes de processar simultaneamente

múltiplas variáveis interdependentes e identificar estruturas latentes nos dados. **(ii) Automatização de Processos Analíticos** (Kanungo et al., 2024), cuja ênfase sistemática na automatização de múltiplas etapas do processo de construção de modelos representa uma rutura paradigmática face às técnicas estatísticas tradicionais, que demandam investimento manual substancial em seleção de variáveis e engenharia de características. Esta automatização revela-se particularmente estratégica em aplicações de *cross-selling*, onde a análise simultânea de múltiplas variáveis — desde dados demográficos e histórico transacional até indicadores comportamentais e preferências manifestas — requer capacidades computacionais que excedem as limitações da intervenção humana manual. **(iii) Orientação Preditiva versus Interpretativa** (van Witteloostuijn et al., 2022), a distinção fundamental entre modelação algorítmica baseada em dados observados, sustentada por validação empírica através de dados de teste e metodologias tradicionais que priorizam a compreensão aprofundada do processo gerador de dados e interpretabilidade causal reflete objetivos epistemológicos divergentes. No contexto específico do *cross-selling* segurador, a supremacia da capacidade preditiva sobre a interpretabilidade detalhada justifica-se pela necessidade operacional de identificar, com precisão estatisticamente significativa, clientes com elevada propensão para aquisição de produtos complementares, independentemente da compreensão causal subjacente aos padrões identificados (Shu & Ye, 2023).

2.4.5. SÍNTESE CRÍTICA E CONVERGÊNCIAS PARADIGMÁTICAS

A transição de abordagens *campaign-oriented* para modelos *customer-centric* (Li et al., 2011) encontra na inteligência artificial um catalisador transformador, capaz de superar as limitações históricas associadas às baixas taxas de conversão (Alijoyo et al., 2025) e à identificação inadequada de oportunidades comerciais (Shah et al., 2012). Essa transformação impulsiona uma convergência entre as estratégias tradicionais de *cross-selling* e as novas capacidades preditivas baseadas em modelos de ML no setor segurador (Akhavan & Hassannayebi, 2024).

Esta convergência manifesta-se através da integração entre os fundamentos relacionais do *cross-selling* – baseados na construção de confiança e proximidade com o cliente (Tolga Akçura & Srinivasan, 2005), bem como cocriação de valor educacional que transcende a transação imediata (Li et al., 2011) – e as capacidades técnicas dos algoritmos de classificação supervisionada (*kNN*, *Logistic Regression*, *Naive Bayes*, *Decision Tree*) sistematizados por (Obi, 2023; Szabó et al., 2024) que, através das três dimensões fundamentais estabelecidas por (Hu et al., 2020) – processamento de grandes volumes de dados, automatização de processos analíticos e orientação preditiva versus interpretativa – permitem operacionalizar a personalização dinâmica defendida por (Verhoef et al., 2009) e a identificação precisa de momentos ótimos para ofertas complementares. O paradigma emergente, validado empiricamente por (Kumar & Reinartz, 2016) na redução de custos de aquisição até 70% e incremento de rentabilidade de 15-25%, e por (Qaraeen et al., 2024) na demonstração da eficácia dos algoritmos preditivos aplicado aos seguros, sugere que o sucesso do *cross-selling* contemporâneo reside não apenas na sofisticação algorítmica, mas na capacidade de harmonizar *insights* comportamentais profundos com precisão

preditiva, transformando dados em conhecimento acionável que beneficia mutuamente ambas as partes através de soluções genuinamente personalizadas e temporalmente adequadas.

3. METODOLOGIA

A presente investigação adota a metodologia CRISP-DM (*Cross-Industry Standard Process for Data Mining*) como *framework* estruturante, reconhecida como *standard* de facto na indústria e academia para projetos de análise preditiva (Chapman et al., 1999; Mariscal et al., 2010), proporcionando um processo iterativo e sistemático que abrange seis fases sequenciais: (i) compreensão do negócio, focada na tradução dos objetivos estratégicos de *cross-selling* da Doutor Finanças em requisitos técnicos específicos para modelos preditivos de propensão de compra de seguros de vida; (ii) compreensão dos dados, caracterizada pela análise exploratória sistemática da base de dados proprietária da MSP, incluindo dados transacionais, comportamentais e demográficos; (iii) seleção e preparação dos dados, tratamento dos valores em falta, balanceamento das classes e normalização, reconhecida como fase crítica que consome tipicamente 80% do esforço total; (iv) modelação, implementação de múltiplos algoritmos de *machine learning* (*kNN*, *Logistic Regression*, *Naive Bayes*, *Tree*); (v) avaliação, análise quantitativa através de métricas de performance como a acurácia (*CA*), *AUC* e análise da *matriz de confusão*; e (vi) implementação, traduz descobertas preditivas em estratégias de personalização comercial, materializando padrões comportamentais identificados em diretrizes operacionais para otimização de *cross-selling* de seguros de vida (Chapman et al., 1999). Este *framework* estabelece um processo metodológico robusto, reproduzível e orientado para resultados, permitindo a transformação do objetivo de personalização da experiência do cliente das MSPs e a otimização das estratégias de *cross-selling* no setor segurador através do desenvolvimento e validação de modelos preditivos do comportamento do consumidor.

3.1. COMPREENSÃO DO NEGÓCIO (*BUSINESS UNDERSTANDING*)

3.1.1. FRAMEWORK ANALÍTICO DAS MSPS: DIMENSÕES ESTRATÉGICAS NO SETOR SEGURADOR

As MSPs relevam uma tipologia multidimensional através de três dimensões analíticas fundamentais que proporcionam uma compreensão holística destes modelos de negócio emergentes, com particular relevância para o setor segurador. A revolução das MSPs transformou modelos de negócio em áreas tradicionais como transportes, hospitalidade e serviços financeiros, permitindo o rápido crescimento de unicórnios, por todo o mundo, com o mercado global a atingir \$234 mil milhões de dólares americanos em 2022 e crescimento esperado para \$2 biliões até 2027 (Teece et al., 2022). Esta transformação económica é comparável à revolução industrial de há 250 anos, uma vez que as plataformas digitais multilaterais remodelam a economia atual através da redução de custos de coordenação entre participantes no mercado, criando e capturando valor diferentemente dos modelos lineares tradicionais (Allweins et al., 2021).

A primeira dimensão, é baseada na classificação mais ampla de modelos de negócio das MSPs, identifica quatro tipos principais que transcendem as fronteiras setoriais e revelam padrões universais de criação e captura de valor. As *transaction platforms* facilitam transações entre diferentes grupos de utilizadores, exemplificadas por eBay, Amazon Marketplace e Uber, criando valor através da facilitação

de trocas e gerando receita através de comissões sobre transações (Mlčúchová, 2022). As **innovation platforms** permitem inovações na plataforma através de terceiros, como *Apple iOS* e *Google Android*, criando valor através da habilitação de inovação externa e gerando receita com o licenciamento e taxas de desenvolvimento. As **integrated platforms** combinam elementos de transação e inovação, exemplificadas por *Amazon (marketplace + AWS)* e *Google (pesquisa + Android)*, diversificando receita através de múltiplos *streams*. As **investment platforms** facilitam investimentos e financiamento, incluindo plataformas de *crowdfunding* e *FinTech*, conectando investidores e beneficiários e gerando receita através de taxas de gestão e sucesso (Mlčúchová, 2022). Os principais atributos identificados nesta dimensão são os dados digitais, fundamentais para a criação de valor através da personalização de serviços, a conectividade e os efeitos de rede, que amplificam atividades de criação de valor onde o valor aumenta com o número de utilizadores, demonstrando a dominância evolutiva sobre modelos tradicionais (Mlčúchová, 2022).

A segunda dimensão, é a classificação funcional, onde organiza as MSP em três categorias principais que refletem diferentes abordagens à facilitação de transações no mercado segurador. As **transaction platforms** facilitam diretamente a compra e venda de seguros, integrando todo o processo desde a cotação até à contratação, incluindo processamento de pagamentos e emissão de apólices, caracterizando-se por elevado nível de integração tecnológica com seguradoras e foco na eficiência do processo de contratação (Micallef et al., 2023). As **non-transaction platforms** concentram-se na comparação e informação sem facilitar diretamente a transação final, focando-se na educação do consumidor, comparação transparente de produtos e preços, e criação de valor através de informação e transparência, com modelo de receita baseado em comissões de referência ou publicidade (Micallef et al., 2023). As **hybrid platforms** combinam elementos de ambas as abordagens anteriores, oferecendo comparação e informação juntamente com capacidades transacionais seletivas, permitindo maior flexibilidade na proposta de valor e adaptação a diferentes tipos de produtos de seguro (Micallef et al., 2023).

A terceira dimensão, é baseada na origem, e oferece *insights* críticos sobre vantagens competitivas estruturais através de uma nova tipologia que classifica as plataformas em três tipos distintos. As **born-platform** representam *startups* nativas digitais cujo modelo de negócio baseia-se numa proposta de valor de plataforma digital desde o início, criadas para preencher nichos particulares e permitir interações entre diversos utilizadores, dando origem a novos mercados (Teece et al., 2022). Estas plataformas caracterizam-se pelo *ADN* nativo de conectividade, com uma arquitetura tecnológica pensada para escalar e criar efeitos de rede, beneficiando de agilidade e flexibilidade tecnológica superior, cultura organizacional orientada para inovação e capacidade de escalar rapidamente. Exemplos paradigmáticos incluem *Uber*, *Airbnb*, *Amazon Marketplace* e redes sociais como *Facebook* e *LinkedIn*. As **platform-born-adjacent** tendem a servir utilizadores finais como *downstream platforms*, invocando inovação transformativa através de diferentes formas organizacionais, endereçando clientes de forma diferente

das *incumbent-born* através de estratégias de *market broadening* que expandem mercados existentes servindo novos segmentos. Representam a expansão natural de negócios existentes para modelos de plataforma, aproveitando ativos e competências já desenvolvidos numa diversificação estratégica que mantém o *core business* original. Esta tipologia beneficia de uma base de clientes estabelecida, recursos e conhecimento de mercado existentes, apresentando menor risco de transição. Exemplos notáveis incluem *Apple* (evolução de *hardware* para *App Store* e serviços), *Microsoft* (transição de *software* para *Azure* e *marketplace*) e *Tesla* (expansão de automóveis para rede de carregamento e *software*) (Teece et al., 2022). As *incumbent-born*, dependentes do produto central do incumbente, tendem a ser *upstream platforms* que enfrentam desafios significativos, tropeçando em configurações organizacionais e capacidades existentes, evidenciando dificuldades em assegurar o sucesso da plataforma digital devido a limitações estruturais. Estas plataformas representam a transformação digital forçada de empresas tradicionais estabelecidas, caracterizando-se pela adaptação às mudanças do mercado digital através de modelos híbridos que integram operações tradicionais com funcionalidades de plataforma. Enfrentam desafios substanciais relacionados com *legacy systems*, resistência à mudança organizacional e competição direta com *players* nativos digitais. Exemplos incluem bancos tradicionais que criam *marketplaces* financeiros, seguradoras que desenvolvem ecossistemas digitais e retalho tradicional que implementa plataformas *omnichannel*. Esta taxonomia revela uma correlação estratégica fundamental onde cada tipo de plataforma adota abordagens de mercado específicas, onde as *born-platforms* exploram oportunidades de criação de mercado através da inovação disruptiva, as *born-adjacent* implementam estratégias de alargamento de mercado aproveitando competências existentes, enquanto as *incumbent-born* focam no aprofundamento de mercado através da digitalização de operações tradicionais. (Teece et al., 2022).

(Pousttchi & Gleiss, 2019) Introduzem uma dimensão estratégica específica para o setor segurador que categoriza as MSPs segundo quatro perfis de interação, revelando diferentes abordagens de posicionamento competitivo no mercado. O perfil **competição** caracteriza-se por plataformas que competem diretamente com seguradoras tradicionais, posicionando-se como alternativas e adotando estratégias competitivas diretas com maior impacto disruptivo. O perfil **coordenação** envolve plataformas que organizam e estruturam a oferta de múltiplas seguradoras, facilitando a comparação e seleção sem competir diretamente, assumindo o papel de coordenador neutro. O perfil **cooperação** manifesta-se através de parcerias estratégicas entre plataformas e seguradoras, criando propostas de valor complementares através de relações de benefício mútuo. O perfil **colaboração** representa o nível mais elevado de integração, com desenvolvimento conjunto de produtos e estratégias de mercado, caracterizada por integração profunda de operações e criação de sinergias transformativas.

3.1.2. COCRIAÇÃO DE VALOR E EFEITOS DE REDE

As MSPs revolucionam a relação tradicional cliente-empresa através da cocriação de valor sistemático ao longo de toda a jornada de compra (McFall et al., 2020; Verhoef et al., 2021). Esta abordagem

contrasta radicalmente com o modelo tradicional onde o cliente é um recetor passivo, transformando-o num participante ativo na criação de valor (Payne & Frow, 2006; Pousttchi & Gleiss, 2019). A cocriação manifesta-se através de múltiplas dimensões, com os utilizadores a contribuírem com feedback contínuo sobre produtos e serviços, participam ativamente no desenvolvimento de novos produtos, partilham experiências que educam outros utilizadores, e colaboram na melhoria contínua da plataforma (Pousttchi & Gleiss, 2019).

A operacionalização de uma orientação *customer-centric* pressupõe a eliminação de silos organizacionais e o alinhamento estratégico transversal focado no cliente. Esta reconfiguração estrutural constitui um fator crítico de sucesso, potenciando a captura de dados multidimensionais através de todos os pontos de contacto. O paradigma de cocriação de valor estabelece uma dinâmica bilateral onde o cliente contribui com dados estratégicos para otimizar a sua própria experiência (Payne & Frow, 2006). Esta lógica evolui nas plataformas digitais para arquiteturas colaborativas onde múltiplos *stakeholders* co-desenvolvem soluções iterativamente (Chan et al., 2022), transcendendo os modelos tradicionais de gestão do conhecimento e criando ecossistemas digitais adaptativos (Hanelt et al., 2021; Payne & Frow, 2006; Pousttchi & Gleiss, 2019).

Esta dinâmica colaborativa estabelece ciclos virtuosos de melhoria contínua, onde o conhecimento gerado pelos utilizadores é sistematicamente capturado, processado e reintegrado na plataforma, criando propostas de valor progressivamente mais sofisticadas e personalizadas. Consequentemente, esta participação ativa não só melhora exponencialmente a qualidade dos serviços através de feedback contínuo e aprendizagem adaptativa, mas também aumenta o *engagement* e a lealdade à plataforma (Brodie et al., 2011), estabelecendo barreiras de saída significativas que consolidam vantagens competitivas sustentáveis (Chan et al., 2022).

Os efeitos de rede constituem o mecanismo fundamental através do qual as MSPs criam e capturam valor, manifestando-se através de cinco tipologias distintas que determinam a dinâmica competitiva e a sustentabilidade das plataformas. ***Efeitos de rede diretos*** (Metcalfe, 1995), onde o valor para um utilizador aumenta com o número de utilizadores do mesmo tipo, exemplificados por redes sociais onde mais utilizadores geram maior conectividade e utilidade. ***Efeitos de rede indiretos*** (Rochet & Tirole, 2006), manifestando-se quando o valor para um tipo de utilizador aumenta com utilizadores de outro tipo, como no caso de sistemas operativos onde mais utilizadores atraem mais desenvolvedores e vice-versa. ***Efeitos de rede de dados*** (Becker et al., 1999), resultantes da melhoria dos serviços através da análise de dados agregados, permitindo a personalização e otimização algorítmica que beneficia todos os utilizadores. ***Efeitos de rede positivos*** (M. L. Katz & Shapiro, 1985), criando ciclos virtuosos de crescimento onde o aumento de participantes gera maior valor, atraindo mais participantes num processo autorreforçado. Contrastando, os ***efeitos de rede negativos*** (Rohlf's, 2003), quando o congestionamento, saturação ou complexidade excessiva reduz o valor percebido pelos utilizadores, podendo comprometer

a viabilidade da plataforma se não forem adequadamente geridos (Mlčúchová, 2022; Pousttchi & Gleiss, 2019; Sanchez-Cartas & León, 2021).

No setor segurador, existem muitas assimetrias de informação ((Méndez-Aparicio et al., 2020; Pousttchi & Gleiss, 2019), os efeitos de rede de dados tornam-se particularmente poderosos (Mlčúchová, 2022; Pousttchi & Gleiss, 2019; Sanchez-Cartas & León, 2021). No ambiente digital, os clientes valorizam qualidade, relevância e clareza da informação, as MSPs utilizam IA para antecipar necessidades e personalizar a experiência da jornada do cliente. Esta capacidade preditiva garante que a qualidade percebida - o determinante importante da experiência de compra (Lemon & Verhoef, 2016; Verhoef et al., 2009) - seja consistentemente superior, com utilidade, informação e tecnologia como fatores críticos diferenciadores (Méndez-Aparicio et al., 2020). Plataformas com maior volume de utilizadores podem oferecer recomendações progressivamente mais precisas e personalizadas, criando ciclos virtuosos de crescimento que tornam praticamente impossível para competidores tradicionais alcançar paridade competitiva (Mlčúchová, 2022).

A utilização de algoritmos de *machine learning* preditivos permite transformar dados em *insights* acionáveis que identificam oportunidades de *cross-selling* inteligente baseadas em padrões comportamentais bidirecionais complexos, proporcionando uma experiência fluida através de múltiplos *touchpoints* e otimização dinâmica das estratégias de interação com o utilizador (Chen & Prentice, 2025).

3.1.3. CONTEXTO ESTRATÉGICO E OPORTUNIDADE DE MERCADO

A Doutor Finanças, enquanto MSP, distingue-se pela sua génese como plataforma nativa digital (*born-platform*), operando através de um modelo de negócio fundamentado na *transaction platform*, com uma dinâmica funcional de uma *hybrid platform* uma vez que integra múltiplas funcionalidades, e adotando um perfil estratégico colaborativo que promove a cocriação de valor entre todos os *stakeholders* do ecossistema segurador português. Atualmente confronta-se com desafios estratégicos significativos na otimização das suas estratégias de *cross-selling*, particularmente no segmento de seguros de vida. A análise da base de clientes revela uma oportunidade estrutural substancial, em que, até final de 2024, 48% dos clientes que possuem pelo menos uma apólice ativa não possuem apólice de seguro de vida, representando um potencial de mercado inexplorado considerável.

Esta lacuna de penetração no segmento de seguros de vida assume particular relevância estratégica quando contextualizada no panorama do mercado segurador português, onde o ramo vida registou uma contração significativa de -12,4% em 2023. A baixa penetração dos seguros de vida na base de clientes da Doutor Finanças espelha os desafios estruturais do mercado nacional, incluindo os reduzidos níveis de poupança e a necessidade de maior literacia financeira dos consumidores.

3.1.4. OBJETIVOS ESTRATÉGICOS PARA O NEGÓCIO DA MODELAÇÃO PREDITIVA

A presente investigação centra-se na melhoria da experiência do cliente no *cross-selling* de seguros de vida junto da carteira de clientes detentores de outras apólices, mediante o desenvolvimento de algoritmos preditivos de suporte a esta estratégia comercial direcionada. Neste contexto, a implementação de abordagens personalizadas, fundamentadas em modelação preditiva, permitirá à Doutor Finanças alcançar múltiplos objetivos estratégicos interconectados ao longo da jornada de compra:

Maximização do valor do cliente (*customer lifetime value*): o aumento da penetração de seguros de vida na base de clientes existente representa uma oportunidade direta de incrementar o valor médio por cliente, aproveitando o relacionamento já estabelecido para introduzir produtos complementares de maior valor agregado.

Otimização da estrutura de custos comerciais: uma das vantagens competitivas fundamentais do *cross-selling* para a base de clientes existente reside na eliminação dos custos de captação de novos clientes. Ao focar na conversão de clientes já estabelecidos, a Doutor Finanças concentra os investimentos comerciais exclusivamente nos custos de conversão, que são significativamente inferiores aos custos de aquisição. Esta otimização da estrutura de custos permite melhorar substancialmente as margens operacionais e o retorno sobre investimento em atividades comerciais.

Personalização temporal e contextual da comunicação: através da identificação precisa da propensão de compra de seguros de vida, torna-se possível implementar estratégias de comunicação altamente personalizadas que impactam cada cliente no momento certo com conteúdo específico e relevante. Esta abordagem substitui comunicações genéricas por comunicações contextualizadas que respondem às necessidades e ao timing específico de cada cliente, maximizando a probabilidade de conversão.

Gestão inteligente da pressão comercial: a modelação preditiva permite implementar uma estratégia diferenciada de abordagem comercial onde clientes com elevada propensão de compra recebem comunicações direcionadas sobre seguros de vida, enquanto clientes com baixa propensão não são "massacrados" com ofertas inadequadas. Para estes últimos, desenvolvem-se abordagens complementares focadas em outros produtos ou em educação financeira preparatória, preservando o relacionamento e preparando futuras oportunidades de conversão.

Fortalecimento do relacionamento cliente-plataforma e redução do churn: a introdução bem-sucedida de seguros de vida, produtos caracteristicamente de longo prazo e elevada importância pessoal, fortalece significativamente os laços relacionais entre cliente e plataforma, criando barreiras naturais à migração para concorrentes e reduzindo substancialmente as taxas de *churn*. Quanto mais duradouro for o relacionamento e maior o número de pontos de contacto estabelecidos, maior será o volume de dados comportamentais recolhidos. Estes dados enriquecem continuamente o conhecimento sobre cada cliente,

permitindo personalizar progressivamente a experiência e criar um ciclo virtuoso de melhoria contínua do relacionamento.

Enriquecimento contínuo da base de dados comportamentais: o aumento dos *touchpoints* de contacto com o cliente, resultante da gestão de múltiplas apólices e da intensificação do relacionamento, gera um fluxo contínuo de dados comportamentais de elevado valor. Estes dados alimentam constantemente os algoritmos preditivos, permitindo melhorar progressivamente a precisão das previsões e otimizar continuamente a experiência do cliente através de recomendações cada vez mais personalizadas e timing de comunicação mais preciso.

Elevação da literacia financeira: o processo de *cross-selling* de seguros de vida constitui uma oportunidade privilegiada para aumentar o nível de educação financeira dos clientes, explicando conceitos fundamentais de proteção financeira, planeamento sucessório e gestão de risco pessoal, contribuindo para decisões mais informadas.

Otimização da satisfação e *engagement*: a oferta proativa de soluções de proteção vida, adequadas às necessidades específicas de cada cliente e apresentadas no timing apropriado, aumenta a perceção de valor da plataforma e demonstra preocupação genuína com o bem-estar financeiro do cliente, elevando os níveis de satisfação e *engagement*.

Consolidação da retenção e fidelização: clientes com múltiplas apólices, particularmente incluindo seguros de vida, apresentam taxas de retenção significativamente superiores, dado o maior *switching cost* e a maior dependência dos serviços da plataforma para gestão do seu portfolio segurador.

Melhoria da performance comercial global: o sucesso nas estratégias de *cross-selling* de seguros de vida impacta diretamente a performance comercial através do aumento das comissões, melhoria dos rácios de conversão, otimização do retorno sobre investimento em atividades de marketing e relacionamento com clientes, e criação de um ciclo virtuoso de crescimento sustentável baseado na maximização do valor da base de clientes existente.

3.2. COMPREENSÃO DOS DADOS (*DATA UNDERSTANDING*)

3.2.1. CARACTERIZAÇÃO DA AMOSTRA E ESTRUTURA DO *DATASET*

O presente estudo fundamenta-se numa amostra representativa de 2415 clientes da Doutor Finanças, estruturada através de um conjunto multidimensional de 37 variáveis que capturam de forma abrangente o perfil e comportamento dos clientes até 2024. A **variável dependente** “*PossuiApoliceVida2024*”, representa uma variável categórica binária (Sim/Não) que determina se o cliente adquiriu uma apólice de seguro de vida durante o ano de 2024, exibe uma distribuição perfeitamente equilibrada, com taxa de conversão de 52% (1260 conversões face a 1155 não conversões), configurando uma característica metodológica de particular relevância. Este equilíbrio distribucional não apenas elimina enviesamentos inerentes a *datasets* desbalanceados (van Witteloostuijn et al., 2022), como também otimiza as condições para a aplicação de técnicas de modelação preditiva, assegurando representatividade

estatística adequada para ambas as classes e proporcionando um ambiente analítico ideal para o desenvolvimento de algoritmos de *machine learning* robustos e com elevada capacidade de generalização.

A estrutura conceptual do *dataset* (anexo: tabela 8) reflete uma abordagem holística de caracterização comportamental, organizando as 37 variáveis por oito categorias principais, proporcionando uma visão multidimensional do perfil e comportamento dos clientes:

- Identificação: 2 variáveis
- Perfil Socioeconómico: 3 variáveis
- Perfil Demográfico: 4 variáveis
- Segmentação Geográfica: 1 variável
- Perfil Tecnológico: 1 variável
- Métricas de *Engagement*: 4 variáveis
- Comportamento Relacional: 9 variáveis
- Carteira de Seguros: 13 variáveis

3.2.2. DIMENSÃO COMPORTAMENTAL E PERFIL SOCIODEMOGRÁFICO

A dimensão comportamental bidirecional constitui uma inovação metodológica distintiva, quantificando através de nove variáveis específicas a dinâmica interacional complexa estabelecida entre o especialista (gestor da Doutor Finanças) e o (paciente) cliente. Esta abordagem captura simultaneamente o comportamento do especialista através de três dimensões críticas - comunicação escrita, capacidade em chamadas telefónicas e disponibilidade temporal - e o comportamento do cliente mediante seis variáveis complementares que incluem comunicação escrita e verbal, capacidade em chamadas, disponibilidade, e tipo de interação preferencial. Esta perspetiva bidirecional permite uma análise holística da qualidade relacional, reconhecendo que a propensão à compra de seguros de vida emerge não apenas das características individuais do cliente, mas da sinergia comunicacional e da compatibilidade comportamental estabelecida entre ambas as partes no processo consultivo.

Esta conceptualização relacional transcende as abordagens convencionais centradas exclusivamente em características demográficas e transacionais, constituindo um diferencial metodológico que permite capturar nuances relacionais determinantes na propensão de compra e representa um contributo significativo para a literatura sobre modelação preditiva no contexto do setor segurador.

A análise descritiva revela um arquétipo de cliente caracterizado por uma predominância da “*Geração X_Jovem*” (87.8%) com competências tecnológicas híbridas, maioritariamente inserida no sector privado (87.8%) e posicionada no segmento de “*rendimento médio-baixo*” (53.1%), apresentando

tipicamente 1-2 dependentes (69.4%) numa distribuição de género perfeitamente equilibrada. A análise da distribuição geográfica evidencia uma marcada concentração urbana, com Lisboa a representar 35% dos registos válidos, seguida pelo Porto (20%) e Coimbra (10%), refletindo o padrão típico de densidade populacional e atividade económica dos principais centros urbanos portugueses. Paralelamente, a distribuição por canais de aquisição revela a supremacia das Leads Referenciadas como estratégia de captação mais eficaz (53.1%), seguidas pelas Leads Não Referenciadas (18.4%), Campanhas Pagas (12.3%) e Marketing Diverso (6,1%), evidenciando a importância crítica do marketing de referência e *word-of-mouth* na geração de leads qualificados, configurando um panorama analítico que combina concentração geográfica urbana com eficácia diferenciada dos canais de aquisição no sector segurador.

3.2.3. PADRÕES COMPORTAMENTAIS

O comportamento relacional evidencia três fenótipos comportamentais distintos – “*Camaleões*” (59.3%) (Perfil híbrido digital/tradicional), “*PoucoDigital*” (22.4%) e “*Digital*” (18.4%) - cada um com implicações estratégicas diferenciadas para a personalização da abordagem comercial e otimização dos *touchpoints* relacionais. A análise às variáveis de *engagement* (anexos: tabela 8) revela padrões de distribuição heterogéneos e elevada dispersão nos dados. A variável *numeroinscriçõesCH* apresenta uma média de 3,42 inscrições por cliente, com mediana inferior (2) e moda unitária (584 ocorrências), indicando uma distribuição assimétrica positiva que se estende até um máximo de 36 inscrições. A variabilidade é substancial, com um desvio padrão de 3,45. A variável *numeroinscriçõesCPCC* exibe um padrão ainda mais disperso, com média de 1,89, moda zero (1.268 ocorrências), refletindo que a maioria dos clientes não possui inscrições neste tipo de serviço. A variabilidade é extremamente elevada (150.8%), com um desvio padrão de 2,85 e amplitude de 18 inscrições. O total agregado de inscrições confirma esta tendência de alta dispersão, apresentando média de 5,31, mediana de 3 e coeficiente de variação de 103%, com uma amplitude máxima de 45 inscrições. Esta elevada variabilidade sugere perfis de utilização de serviços marcadamente diferenciados.

3.3. PREPARAÇÃO DOS DADOS (DATA PREPARATION)

3.3.1. ESTRATÉGIA DE TRATAMENTO DAS VARIÁVEIS

Considerando a natureza dos dados empresariais e os objetivos de interpretabilidade do modelo, foi implementada uma estratégia de conversão de variáveis numéricas específicas em variáveis categóricas através da criação de intervalos significativos do ponto de vista empresarial. As variáveis de comportamento, originalmente numa escala numérica, foram categorizadas em intervalos ordenados que refletem padrões comportamentais distintos. Esta abordagem alinha-se com a prática recomendada de discretização de variáveis contínuas em categorias ordinais para facilitar a interpretação dos dados e preservem informação relevante sobre a hierarquia comportamental. Os rendimentos estão agrupados em intervalos que refletem segmentos de mercado relevantes para a Doutor Finança.

3.3.2. CARACTERIZAÇÃO ESTRUTURAL DO DATASET

A análise confirma a integridade estrutural do *dataset* com 2.415 registos e 37 variáveis, apresentando uma arquitetura de dados bem definida com a variável target “*PossuiApoliceVida2024*” (Sim/Não) e identificador único “*PatientID*”. A tipologia das variáveis revela uma composição diversificada e equilibrada, entre variáveis categóricas nominais representando a maioria dos atributos, complementadas por variáveis numéricas (“*NumeroInscricoesCH*”, “*NumeroInscricoesCPCC*” e “*NºDependentes*”) e variáveis binárias correspondentes ao histórico de transações por tipo de seguros. Esta distribuição tipológica proporciona uma base sólida para modelação preditiva, combinando informação demográfica, comportamental e transacional. Garantindo flexibilidade metodológica na abordagem ao problema de classificação binária da propensão de compra de seguros de vida.

3.3.3. AVALIAÇÃO DA QUALIDADE DOS DADOS

O *dataset* apresenta ausência de valores omissos, com 100% de completude nas variáveis de comportamento relacional, nas variáveis da carteira de seguros, e nas variáveis de medem o *engagement* (anexo: tabela 8). A distribuição dos segmentos de rendimento revela uma estrutura, com predominância das classes média-baixa (45%) e média (32%), seguidas pelas classes média-alta (12%), baixa (8%) e premium (3%). O histórico das variáveis de carteira de seguros mantém consistência metodológica entre o *baseline* anterior a 2024 e o período de 2024, proporcionando uma base sólida para avaliar a evolução comportamental da carteira e estabelecer padrões preditivos baseados na trajetória temporal dos dados. Esta completude nas dimensões analíticas centrais compensa parcialmente as limitações identificadas em variáveis sociodemográficas secundárias, configurando um panorama favorável para o desenvolvimento de modelos preditivos eficazes e estratégias de *cross-selling* personalizadas baseadas nos perfis comportamentais e histórico de relacionamento com a organização.

3.3.4. IDENTIFICAÇÃO E ANÁLISE DE LIMITAÇÕES

As limitações identificadas e que exigem tratamento metodológico rigoroso antes da modelação, nomeadamente inconsistências lógicas críticas onde 866 registos (35.9%) apresentam contradições entre variáveis de seguro (“*SemSeguroAntes2024 = Sim*” coexistindo com “*PossuiApoliceSaudeAntes2024 = Sim*”), problemas de *encoding* inconsistente em variáveis de comunicação, e redundância confirmada entre “*SegmentEtárioGeração*” e “*SegmentCompTecnologico*”. Adicionalmente, a variável “*Distrito*” apresenta 32.4% de valores em falta, enquanto um cluster de variáveis sociodemográficas exhibe taxas de *missing data* (Kapoor et al., 2025) entre 23.1% e 25.7%, criando potencial para viés geográfico e demográfico.

3.3.5. IMPLEMENTAÇÃO DO TRATAMENTO DE DADOS

No tratamento dos dados foi tratada a correção das inconsistências lógicas através de regras de negócio validadas, padronização do *encoding* de caracteres, remoção da redundância entre variáveis geracionais e tecnológicas, retirando do modelo a variável “*SegmentCompTecnologico*”, e implementação de *feature engineering* estratégico para criar a variável “*TotalInscrições*” que representa o numero total de

inscrições por cliente nos diferentes produtos disponibilizados pelo Doutor Finanças (SEGUROS, CH, CCPC). As contradições críticas identificadas em 866 registos (35.9%), entre "SemSeguroAntes2024" e "PossuiApoliceSaudeAntes2024" foram tratadas através de regras de negócio específicas, onde a presença de apólice de saúde prevalece sobre a declaração genérica de ausência de seguro, seguindo as melhores práticas de *data cleaning* (Ganti & Sarma, 2013) que recomendam a resolução de conflitos através de hierarquização lógica. Para o tratamento do *missing data*, a imputação múltipla é geralmente a abordagem mais robusta e flexível, permitindo análises estatísticas padrão enquanto preserva a incerteza associada aos valores em falta (A. Kumar et al., 2025). Foi tomada a decisão estratégica de não implementar o tratamento para o *missing data*, optando por preservar a estrutura original dos dados (Kapoor et al., 2025).

3.4. MODELAÇÃO (*MODELING*)

3.4.1. PLATAFORMA DE IMPLEMENTAÇÃO

A implementação da estratégia de modelação foi concretizada através da plataforma Orange versão: 3.38.1, uma ferramenta que se distingue pela sua arquitetura visual baseada em *widgets* modulares. Esta abordagem representa uma evolução significativa na forma como os processos *de machine learning* são conceptualizados e executados, oferecendo uma alternativa transparente e intuitiva aos métodos tradicionais de programação. Esta metodologia alinha-se perfeitamente com os princípios de transparência e reprodutibilidade exigidos na investigação científica contemporânea.

3.4.2. ARQUITETURA DO WORKFLOW DE DADOS

Foi elaborada uma divisão estratificada 70% *train* - 30% *test* do *dataset*. A escolha da proporção 70%-30% baseia-se numa análise cuidadosa da adequação estatística para o *dataset* em questão (Breiman, 2001). Num total de 2.415 observações no *dataset*, o conjunto de treino terá 1691 observações e o conjunto de teste terá 724 observações. O conjunto de treino proporciona dados suficientes para treino robusto dos quatro algoritmos, enquanto o conjunto de teste mantém poder estatístico adequado para uma validação independente (Nagaraju et al., 2023). A variável *target* (Ka et al., 2023) definida é "*PossuiApoliceVida2024*", servindo de base para a construção dos modelos preditivos de *cross-selling*.

3.4.3. ESTRATÉGIA DE MODELAÇÃO: ALGORITMOS PARA CROSS-SELLING

Os algoritmos identificados na literatura como sendo muito utilizados no setor segurador incluem *k-Nearest Neighbors (kNN)*, *Logistic Regression*, *Naive Bayes*, e *Decision Tree* (Obi, 2023; Szabó et al., 2024).

***K-Nearest Neighbors (kNN)*:** Este algoritmo classifica novas instâncias baseando-se na classe maioritária dos seus *k* vizinhos mais próximos no espaço de características. No contexto do *cross-selling* segurador, o KNN é particularmente relevante porque assume que clientes com perfis similares tendem a exibir comportamentos de compra similares (Lim et al., 2023; Szabó et al., 2024). As vantagens incluem simplicidade conceptual, capacidade de capturar padrões locais complexos e robustez a *outliers*.

As limitações compreendem sensibilidade à dimensionalidade e intensidade computacional para grandes conjuntos de dados (Mahadevkar et al., 2022).

Logistic Regression: Constitui um método estatístico clássico que modela a probabilidade de pertença a uma classe utilizando a função logística (Obi, 2023; Szabó et al., 2024). No *cross-selling*, permite estimar a probabilidade de um cliente adquirir um produto específico baseado nas suas características. A principal vantagem reside na interpretabilidade direta dos coeficientes, permitindo compreender quais os fatores mais influenciam as decisões de compra (Hu et al., 2020).

Naive Bayes: Baseia-se no teorema de Bayes com pressuposto de independência condicional entre variáveis preditivas. Apesar desta assunção frequentemente não se verificar na realidade, o algoritmo demonstra um desempenho surpreendentemente bom em muitas aplicações do mundo real, incluindo predição de comportamento do cliente. É particularmente eficaz com múltiplas variáveis categóricas e conjuntos de dados relativamente pequenos (Domingos & Pazzani, 1997; Qazi et al., 2020).

Tree (designação atribuída pelo *Orange*): Criam modelos através da partição recursiva de dados em grupos progressivamente mais homogêneos baseada em critérios especificados até que seja atingida uma regra de paragem. A principal vantagem é a interpretabilidade excepcional, permitindo compreensão clara do processo de decisão e capacidade de identificar regras específicas que caracterizam clientes com alta propensão para produtos específicos (De Caigny et al., 2018; Hu et al., 2020).

Numa perspetiva académica, estes modelos têm contribuído significativamente para a compreensão do comportamento do cliente no setor segurador. No estudo de Tasneem Qaraeen et al., 2024, a probabilidade de comprar um seguro de viagem utilizando múltiplos algoritmos revelou que o *kNN* inesperadamente superou outros modelos com um *CA* de 0,81 e uma *AUC* de 0,78. Este resultado é particularmente interessante porque *kNN* superou algoritmos mais complexos como *Random Forest* e *SVM*, sugerindo que a simplicidade algorítmica pode ser vantajosa em determinados contextos do setor segurador. O desempenho do *kNN* pode ser atribuído à natureza dos dados de seguros, onde clientes com perfis demográficos e comportamentais similares tendem a exibir padrões de compra similares.

3.4.4. MÉTRICAS DE AVALIAÇÃO

A avaliação rigorosa dos algoritmos preditivos é essencial para garantir a sua eficácia em aplicações práticas de *cross-selling* (Szabó et al., 2024). As métricas principais identificadas na literatura incluem a acurácia (*CA*) e *AUC*, cada uma fornecendo perspetivas complementares sobre o desempenho do modelo (Obi, 2023).

Accuracy (CA): Representa a proporção de predições corretas sobre o total de predições feitas, constituindo uma métrica intuitiva calculada como $(\text{Verdadeiros Positivos} + \text{Verdadeiros Negativos}) / \text{Total de Casos}$. No contexto do *cross-selling*, uma *CA* de 0,7 significa que o modelo prevê corretamente 70% dos casos de clientes que irão comprar seguros de vida (Chicco & Jurman, 2020). Quanto maior for a *CA*, melhor será a capacidade do modelo de prever corretamente (Zaki & Meira Jr., 2014).

Embora a *CA* seja de interpretação intuitiva, revela-se altamente problemática em conjuntos de dados desequilibrados (van Witteloostuijn et al., 2022), uma vez que os modelos podem alcançar uma elevada precisão simplesmente prevendo a classe majoritária, podendo ocultar deficiências na identificação de classes minoritárias, mas relevantes (Luo et al., 2009).

Área Sob a Curva ROC (AUC): Resulta da curva *Receiver Operating Characteristic (ROC)*, que representa graficamente a taxa de verdadeiros positivos contra a taxa de falsos positivos em diferentes limiares de decisão, em que a área total do gráfico é 1, a AUC situa-se no intervalo [0,1] e quanto maior, melhor o desempenho. Valores próximos de 0,5 indicam um desempenho equivalente ao acaso (Powers, 2011). No *cross-selling*, um AUC acima de 0,8 indica que o modelo pode efetivamente distinguir entre clientes com alta e baixa propensão para adquirir produtos adicionais.

Tendo uma estrutura de dados equilibrada, procuramos uma AUC elevada, porque significa que o modelo consegue distinguir com elevada precisão as classes positivas e negativas (Obi, 2023). Contudo, a AUC apresenta limitações significativas quando aplicada a conjuntos de dados desequilibrados, uma vez que pode não refletir com precisão o desempenho do modelo quando uma classe supera significativamente a outra em termos de representatividade, e nestas circunstâncias, recomenda-se a utilização de métricas complementares para uma avaliação mais robusta (Saito & Rehmsmeier, 2015).

A interpretação e visualização de métricas de avaliação para modelos de classificação supervisionada requerem uma consideração criteriosa das características do conjunto de dados e dos requisitos específicos do problema. Métricas como a AUC e a CA são amplamente utilizadas, contudo é fundamental ter consciência das suas limitações, particularmente em conjuntos de dados desequilibrados. A compreensão dos pontos fortes e fracos de cada métrica permite desenvolver um espírito crítico sobre os resultados obtidos por cada modelo, possibilitando uma visão mais aprofundada para a tomada de decisões informadas (Chicco & Jurman, 2020; M.N. & Hossin, 2015).

3.4.5. MATRIZ DE CONFUSÃO

Constitui uma ferramenta valiosa para analisar o desempenho de um modelo de previsão de classificação supervisionada (Zaki & Meira Jr., 2014). A sua utilidade é particularmente proeminente no contexto de problemas de classificação binária, onde o foco incide sobre duas classes distintas: classificação positiva e classificação negativa. Permite um exame detalhado das previsões do modelo, desagregando os resultados em verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. Tal granularidade auxilia numa compreensão abrangente dos pontos fortes e fracos do modelo, fornecendo perspetivas sobre a sua capacidade de classificar corretamente as instâncias e identificar áreas de melhoria no processo de predição (Obi, 2023; Saito & Rehmsmeier, 2015; Szabó et al., 2024).

Esta fundamenta-se em quatro categorias essenciais que permitem avaliar a performance de algoritmos preditivos. Os Verdadeiros Positivos representam instâncias onde o modelo identifica

corretamente clientes que adquirirão um produto, como um seguro de vida. Os Verdadeiros Negativos correspondem à identificação precisa de clientes que não efetuarão a compra. Por oposição, os Falsos Positivos ocorrem quando o modelo prevê incorretamente que um cliente adquirirá o produto quando tal não se verificará, enquanto os Falsos Negativos manifestam-se na situação inversa, onde o modelo falha em identificar clientes com real propensão de compra. A análise da matriz de confusão constitui uma base fundamental para a avaliação dos resultados dos modelos de classificação, permitindo definir estratégias de *cross-selling* inteligente no setor segurador (M.N. & Hossin, 2015; N. Srivastava et al., 2014).

3.4.6. FRAMEWORK DE TREINO E TESTES

O processo de construção implementa o treino simultâneo de todos os algoritmos através de paralelização que aproveita otimamente a capacidade computacional disponível, assegurando que todos os modelos treinem no mesmo conjunto de dados para garantir comparação justa e consistência metodológica rigorosa. O processo de validação comparativa submete uniformemente todos os algoritmos ao conjunto de dados teste reservado (30% dos dados), estabelecendo um protocolo metodológico rigoroso que assegura imparcialidade na avaliação. A análise de performance culmina num ranking fundamentado dos modelos baseado em métricas relevantes ao contexto de negócio, incorporando análise sistemática de erros para identificação de padrões nos casos mal classificados, enquanto a seleção final pondera simultaneamente performance quantitativa, interpretabilidade e requisitos específicos do negócio, assegurando alinhamento estratégico com objetivos organizacionais.

As seleções das quatro técnicas de modelação implementadas refletem uma estratégia deliberada de diversificação de paradigmas algorítmicos, assegurando uma cobertura abrangente do espaço de soluções possíveis para o problema de propensão a seguros de vida (De Caigny et al., 2018).

O algoritmo *Logistic Regression* constitui um modelo linear generalizado que oferece coeficientes interpretáveis através de *odds ratios* e probabilidades naturalmente calibradas, estabelecendo um *baseline* robusto para comparação, particularmente adequado quando as relações entre variáveis são aproximadamente lineares. A configuração implementada utilizou regularização *Ridge* (L2) com parâmetro $C=1$, que estabelece um nível médio de regularização estatisticamente equilibrado. A regularização L2 introduz uma penalização baseada na soma dos quadrados dos coeficientes, prevenindo eficazmente o *overfitting* (Wang & Raj, 2017) através do encorajamento de pesos menores e mais distribuídos no modelo. O valor $C=1$ constitui uma *baseline* metodológica amplamente adotada. Esta configuração estabelece um compromisso otimizado entre a capacidade de ajuste aos dados de treino e a manutenção de uma boa generalização para dados não observados, sendo apropriada para a maioria dos *datasets* sem necessidade de ajuste específico. A opção “*Balance class distribution*” foi deliberadamente desativada nesta implementação. Quando ativada, esta funcionalidade ajusta automaticamente o algoritmo para atribuir maior importância às classes minoritárias através de

ponderação inversa à frequência, melhorando significativamente a performance em problemas de classificação caracterizados por distribuições desiguais entre classes (Shu & Ye, 2023).

O algoritmo *Naive Bayes* fundamenta-se rigorosamente no teorema de *Bayes*, demonstrando eficácia particular com variáveis categóricas e robustez notável com *datasets* de dimensão moderada, proporcionando treino e predição computacionalmente eficientes sob o pressuposto simplificador de independência condicional entre *features* (Zaki & Meira Jr., 2014).

O algoritmo *Tree*, utiliza uma estrutura hierárquica baseada em regras, garantindo alta interpretabilidade por meio de regras *if-then* simples e compreensíveis para *stakeholders* não técnicos, capturando automaticamente interações não-lineares sem requerer normalização de dados, embora apresentem tendência ao *overfitting*. Os parâmetros configurados incluíram: um valor mínimo de 2 instâncias por folha terminal e a restrição de não dividir subconjuntos com menos de 3 observações, estabelecendo limites estatisticamente fundamentados que asseguram a robustez das divisões e previnem a criação de nós com representatividade insuficiente. A limitação da profundidade máxima a 3 níveis foi implementada como estratégia de controlo da complexidade, reduzindo significativamente o custo computacional e melhorando a capacidade de generalização do modelo. O critério de paragem definido quando a maioria das observações atinge 95% de pureza representa um *threshold* otimizado que equilibra precisão preditiva e simplicidade estrutural, constituindo uma abordagem de regularização preventiva eficaz contra *overfitting* e garantindo a estabilidade do modelo em dados não observados.

Finalmente, o *kNN* implementa uma aprendizagem baseada em instâncias, classificando mediante a maioria dos *k* vizinhos mais próximos e oferecendo uma abordagem não-paramétrica que se adapta a fronteiras de decisão complexas, não obstante a sensibilidade à escala das variáveis e à dimensionalidade. A implementação do algoritmo *kNN* foi configurada com parâmetros específicos que refletem uma abordagem metodológica equilibrada e fundamentada. A seleção da distância *Euclidiana* (Soofi & Awan, 2017) como métrica representa a escolha mais comum em algoritmos *kNN*, calculando a "linha reta" entre pontos no espaço multidimensional, sendo particularmente eficaz quando as variáveis possuem escalas similares e as relações entre características são predominantemente lineares. O sistema de ponderação uniforme adotado significa que todos os *k=5* vizinhos mais próximos contribuem igualmente para a predição final, independentemente da sua distância exata ao ponto de consulta, contrastando com esquemas de ponderação por distância onde vizinhos mais próximos recebem maior influência na decisão. A escolha de *k=5* vizinhos representa um compromisso equilibrado entre *bias* e variância no modelo, evitando tanto o risco de *overfitting* associado a valores baixos de *k* (*k=1,3*) quanto o *over-smoothing* (Shu & Ye, 2023) resultante de valores elevados, proporcionando robustez suficiente contra ruído enquanto mantém sensibilidade a estruturas locais dos dados, configurando-se como uma *baseline* metodológica sólida para estudos comparativos (Lim et al., 2023).

3.5. AVALIAÇÃO (EVALUATION)

3.5.1. INTRODUÇÃO SINTÉTICA

A análise comparativa entre as fases de treino e teste constitui um elemento fundamental para avaliar a capacidade de generalização dos modelos de *machine learning* implementados, permitindo identificar padrões de degradação de performance e estabilidade preditiva. Esta avaliação metodológica, baseada na comparação sistemática das métricas AUC e CA entre ambientes controlados e cenários reais, revela dinâmicas comportamentais essenciais para a seleção do modelo mais adequado à implementação prática do sistema de recomendação.

3.5.2. ANÁLISE DE PERFORMANCE GLOBAL

A análise comparativa entre os resultados de treino e teste revela padrões fundamentais sobre a capacidade de generalização dos modelos implementados, demonstrando uma degradação controlada da performance. Durante a fase de treino, o algoritmo *kNN* evidenciou supremacia discriminativa com AUC de 0,978, enquanto o *Logistic Regression* demonstrou excelência preditiva superior, alcançando CA de 0,923, revelando assim perfis complementares de otimização entre capacidade de separação de classes e precisão classificatória global. Em contraste, durante a fase de teste, o modelo *Logistic Regression* consolidou a sua robustez discriminativa com AUC de 0,942, enquanto o *Naive Bayes* evidenciou otimização classificatória com CA de 0,906, revelando assim uma redistribuição hierárquica de performance que reflete a capacidade diferencial de generalização dos algoritmos. O modelo *kNN* que, apesar da sua anterior dominância em treino, registou a performance mais conservadora em teste com AUC de 0.922 e CA de 0.894 (tabela 1 e tabela 2).

Nesta primeira análise, podemos concluir que modelos mais simples e baseados em princípios estatísticos sólidos tendem a ser mais confiáveis em situações reais, onde os dados podem diferir ligeiramente dos dados de treino. A simplicidade e robustez estatística muitas vezes superam a complexidade excessiva.

3.5.3. ANÁLISE DE *OVERFITTING*

Numa segunda fase de análise aos resultados, confirma-se a existência de *overfitting* em alguns modelos, manifestando-se com diferentes graus de severidade (anexo: tabela 3). A análise comparativa revela diferentes padrões de generalização entre os algoritmos testados, onde o gap entre performance de treino e teste serve como indicador fundamental de *overfitting* (Geman et al., 1992). Quanto maior for a diferença, maior a probabilidade de o modelo ter memorizado os dados de treino em vez de aprender padrões verdadeiramente generalizáveis. Esta métrica permite identificar modelos que apresentam ajuste excessivo e orientar estratégias de otimização adequadas (Hastie et al., 2004).

Os algoritmo *kNN* demonstra o maior *overfitting*, com quedas significativas de -0,056 no AUC (0,978 → 0,922) e -0,016 na CA (0,910 → 0,894). Este comportamento é típico do *kNN* com valores baixos de k, que tende a memorizar ruído nos dados de treino (anexo: tabela 3). A análise do desempenho do algoritmo *kNN* com K=5 revela um cenário de *overfitting* moderado, onde a queda de -0,056 no AUC

é considerável, mas não excepcional para esta configuração, indicando que o modelo está a capturar padrões nos dados de treino que não se generalizam adequadamente ao conjunto de teste. Curiosamente, a queda menor na CA (-0,016) sugere que, apesar do *overfitting* na capacidade discriminativa medida pelo AUC, a precisão geral da classificação mantém-se relativamente estável. Comparativamente aos outros modelos analisados, o *kNN* apresenta o maior gap de performance, contrastando significativamente com a estabilidade demonstrada pelo Naive Bayes (-0,004 AUC), o que confirma que $K=5$ permite flexibilidade excessiva para este *dataset* específico.

O algoritmo *Tree* apresenta *overfitting* moderado, com queda de -0,033 no AUC (0,957 \rightarrow 0,924) e mínima de -0,006 na CA (0,912 \rightarrow 0,906), possivelmente devido à excessiva ramificação dos nós. O desenvolvimento de divisões hiperespecíficas durante o treino, criou nós terminais excessivamente puros que capturam ruído em vez de padrões genuínos, resultando em regras de decisão que se ajustam perfeitamente ao conjunto de treino, mas geram fronteiras de decisão irregulares e fragmentadas que comprometem a capacidade de generalização para novos dados (Blockeel et al., 2023). A análise dos resultados do algoritmo *Tree* revela um padrão de *overfitting*, possivelmente devido à excessiva ramificação dos nós durante a construção da árvore. O desenvolvimento de divisões hiperespecíficas no processo de treino gerou nós terminais excessivamente puros que capturam ruído em detrimento de padrões genuínos, resultando em regras de decisão que se ajustam perfeitamente ao conjunto de treino, mas geram fronteiras de decisão irregulares e fragmentadas que comprometem a capacidade de generalização para novos dados (De Caigny et al., 2018). A disparidade entre a queda no AUC (-0,033) e na CA (-0,006) ocorre porque o AUC avalia a capacidade discriminativa através de todos os *thresholds* possíveis, sendo extremamente sensível à ordenação incorreta de probabilidades entre classes, à calibração deficiente das probabilidades preditas e às variações na confiança das predições, enquanto a CA foca exclusivamente na decisão final (classe predita), mantendo-se mais robusta a pequenas flutuações probabilísticas desde que a classificação da classe majoritária permaneça correta.

O algoritmo *Logistic Regression* exhibe *overfitting* moderado, com quedas de -0,026 no AUC (0,968 \rightarrow 0,942) e -0,016 na CA (0,918 \rightarrow 0,902). Este modelo linear demonstra maior resistência ao *overfitting* devido à sua simplicidade intrínseca, mantendo boa capacidade de generalização. Esta performance estável reflete as características intrínsecas dos modelos lineares e os seus mecanismos de regularização natural. A simplicidade estrutural da *Logistic Regression* funciona como uma forma implícita de regularização. Ao contrário de modelos de maior complexidade, este está intrinsecamente limitado a aprender apenas relações lineares entre as *features* e a variável target, impedindo eficazmente a captura de padrões artificiais ou ruído específico do conjunto de treino. Esta restrição de capacidade modelar cria um viés indutivo forte que privilegia soluções parcimoniosas e generalizáveis, resultando numa degradação mínima da performance entre os conjuntos de treino e teste (Shu & Ye, 2023).

O *Naive Bayes* destaca-se com o menor *overfitting* observado, apresentando uma queda mínima de -0,004 no AUC (0,939 \rightarrow 0,935) e -0,020 na CA (0,923 \rightarrow 0,903), demonstrando a melhor estabilidade

entre os modelos avaliados, particularmente na métrica AUC. Esta robustez excepcional deve-se fundamentalmente às suposições de independência condicional entre *features*, que funcionam como uma forma natural de regularização. Esta restrição estrutural impede que o modelo se ajuste excessivamente aos padrões específicos dos dados de treino, limitando a sua capacidade de memorizar ruído ou características específicas do conjunto de treino. A diferença quase negligível no AUC (-0,004) revela que o modelo mantém praticamente inalterada a sua capacidade discriminativa entre os conjuntos de treino e teste, enquanto a ligeira queda na CA (-0,020) permanece dentro de limites aceitáveis. Esta consistência notável indica uma excelente capacidade de generalização, sugerindo que o *Naive Bayes* consegue capturar eficazmente os padrões subjacentes verdadeiros do *dataset* sem se deixar influenciar por correlações fortuitas ou artefactos específicos dos dados de treino (Aravind et al., 2023).

Como conclusão podemos estabelecer uma hierarquia clara de suscetibilidade ao *overfitting* entre os algoritmos testados, com o *kNN* a apresentar o maior *overfitting*, seguido pelo *Trees* e *Logistic Regression*, enquanto o *Naive Bayes* demonstra a menor suscetibilidade. Esta hierarquia correlaciona-se diretamente com a complexidade e flexibilidade de cada algoritmo, sendo que os modelos com melhor resistência ao *overfitting* incorporam mecanismos de regularização implícitos. O *Naive Bayes* através das suposições de independência condicional que limitam a memorização de ruído específico, e a *Logistic Regression* mediante a sua simplicidade estrutural linear que impede a captura de padrões artificiais. O gap entre a performance de treino e teste emerge como métrica fundamental para identificar *overfitting*, permitindo distinguir modelos que memorizaram os dados de treino daqueles que aprenderam padrões verdadeiramente generalizáveis, constituindo assim uma ferramenta essencial para a seleção de modelos em aplicações práticas (Aravind et al., 2023; Soofi & Awan, 2017).

3.5.4. PERFIS COMPORTAMENTAIS DOS MODELOS

Numa avaliação granular aos modelos por métricas específicas - AUC e CA - permite identificar padrões distintos, evidenciando que a excelência em treino não se traduz necessariamente em robustez de generalização, enquanto modelos com performance moderada inicial podem demonstrar superior estabilidade preditiva em cenários reais.

AUC (Capacidade de Discriminação):

- *kNN* tem o melhor AUC no treino (0,978), porém, sofre a maior degradação no teste real (-0,056);
- *Naive Bayes* mantém AUC mais estável na passagem entre treino para o teste real;
- *Logistic Regression* consegue atingir um bom desempenho e estabilidade, em treino e no teste real;
- *Tree* sofre uma quebra considerável entre o treino e o teste real (-0,033);

CA (Precisão Global):

- *Tree* é o mais estável de todos os modelos na passagem para o teste real (-0,006) e apesar de no treino ser apenas o 3º melhor resultado, em teste real passa para 1º lugar;
- *Logistic Regression* é bastante estável;
- *kNN* apesar de só ter tido uma quebra reduzida (-0,016) entre treino e teste real, fez com que o resultado seja inferior a 0.9, sendo o pior classificado entre os modelos.
- *Naive Bayes* é quem sobre a maior quebra no CA (-0,02);

A análise comparativa dos perfis comportamentais com implicações práticas específicas:

Tabela 4: Ranking geral (AUC + CA)

Modelo	Estabilidade AUC	Estabilidade CA	Ranking Geral
Tree	3º (-0,033)	1º (-0,006)	1º
Logistic Regression	2º (-0,026)	2º (-0,016)	2º
Naive Bayes	1º (-0,004)	4º (-0,020)	3º
kNN	4º (-0,056)	2º (-0,016)	4º

O *Tree* emerge como a solução mais robusta, demonstrando consistência superior em CA com a menor degradação entre fases (-0,006), tornando-se ideal **devido à sua explicabilidade** (Blockeel et al., 2023), embora apresente o *trade-off* de AUC que sofre uma quebra considerável (-0,033) e como já evidenciamos anteriormente tem alguns desafios de *overfitting* ligeiro, mantendo ainda assim performance discriminativa acima dos 0,92.

O *Naive Bayes* posiciona-se como **especialista em capacidade discriminativa** (Prasetyowati & Sibaroni, 2024), exibindo estabilidade excepcional no AUC durante a transição treino-teste, mas revelando maior variabilidade em precisão global (CA), configurando-se como solução ótima quando a capacidade de ranking constitui prioridade analítica.

O algoritmo *Logistic Regression* estabelece-se como a **alternativa equilibrada** (Ka et al., 2023), alcançando consistentemente a 2º melhor performance em ambas as métricas, evidenciando versatilidade que a torna aplicável a múltiplos cenários quando se procura um compromisso entre discriminação e precisão classificatória.

O *kNN* apresenta um *overfitting* significativo, com degradações consideráveis na performance entre treino e teste, o que compromete gravemente a sua capacidade de generalização. Esta instabilidade inviabiliza a sua utilização em cenários reais onde a robustez e consistência são fundamentais.

3.5.5. ANÁLISE GRANULAR DAS MATRIZES DE CONFUSÃO

A análise da matriz de confusão permite uma avaliação granular da performance classificatória, revelando capacidades específicas de cada modelo na identificação correta de ambas as classes,

proporcionando insights fundamentais sobre pontos fortes e limitações discriminativas individuais (Anexos - Imagem 1, Imagem 2, Imagem 3, Imagem 4).

Os resultados de Verdadeiros Negativos evidenciam a supremacia da *Logistic Regression* (86%) na identificação correta da classe negativa, enquanto o *Naive Bayes* (81.5%) apresenta a menor capacidade de reconhecimento de casos negativos, sugerindo uma tendência para classificações falso-positivas que pode comprometer a especificidade do modelo em aplicações onde a precisão na identificação de "Não" constitui requisito crítico.

Verdadeiros negativos (*true negative rate*)

Capacidade de identificar corretamente os "Não"

1. *Logistic Regression*: 86% (melhor)
2. *Tree*: 84.6%
3. *kNN*: 84%
4. *Naive Bayes*: 81.5%

A capacidade de identificação correta dos casos positivos revela um cenário inverso ao observado nos verdadeiros negativos, com o *Naive Bayes* a demonstrar uma performance excepcional (99.5%), praticamente eliminando falsos negativos e evidenciando uma sensibilidade superior na detecção da classe "Sim". O *Logistic Regression* (94.3%) apresentou o pior desempenho. Esta distribuição sugere que o *Naive Bayes* possui uma tendência natural para classificações positivas, o que explica simultaneamente a sua excelência em *True Positive Rate* e a sua limitação em *True Negative Rate*, caracterizando um perfil algorítmico otimista que privilegia a captura de casos positivos, mesmo à custa de algum incremento em falsos positivos.

Verdadeiros positivos (*true positive rate*)

Capacidade de identificar corretamente os "Sim"

1. *Naive Bayes*: 99.5% (melhor)
2. *Tree*: 95.9%
3. *kNN*: 94.6%
4. *Logistic Regression*: 94.3%

A taxa de falsos positivos confirma o padrão comportamental previamente identificado, com o *Logistic Regression* apresentando o menor índice de recomendações incorretas (14%), demonstrando uma maior precisão na contenção de sugestões inadequadas, enquanto o *Naive Bayes* registra a maior taxa de falsos positivos (18.5%), corroborando a sua tendência otimista para classificações positivas.

Falsos positivos

Produtos recomendados incorretamente

1. *Logistic Regression*: 14% (melhor)
2. *Tree*: 15.4%
3. *kNN*: 16%
4. *Naive Bayes*: 18.5% (Pior)

A análise dos falsos negativos revela o *Naive Bayes* como modelo excepcional na minimização de oportunidades perdidas (0.5%), maximizando assim o potencial de conversão e satisfação do cliente. Em contraste, o *Logistic Regression* apresenta a maior taxa de falsos negativos (5.7%).

Falsos negativos

Oportunidades perdidas de recomendação

1. *Naive Bayes*: 0.5% (Melhor)
2. *Tree*: 4.1%
3. *kNN*: 5.4%
4. *Logistic Regression*: 5.7% (Pior)

A análise comparativa entre as fases de treino e teste revela padrões fundamentais que transcendem métricas isoladas, evidenciando que a excelência em ambiente controlado não se traduz necessariamente em robustez operacional. Os resultados demonstram que modelos baseados em princípios estatísticos sólidos e mecanismos de regularização natural tendem a preservar melhor a performance em cenários reais, sugerindo que a simplicidade algorítmica pode constituir uma vantagem competitiva face à complexidade excessiva. Esta análise estabelece as bases empíricas para a seleção do modelo final, considerando não apenas métricas de performance isoladas, mas perfis comportamentais integrados que reflitam capacidades de generalização e estabilidade preditiva.

3.6. CONCLUSÕES (DEPLOYMENT)

3.6.1. VALIDAÇÃO EMPÍRICA E SÍNTESE DOS RESULTADOS

A presente investigação culmina com uma análise integrada dos resultados obtidos através da implementação de múltiplas abordagens de *machine learning* para a predição de propensão a seguros de vida em contexto de *cross-selling*. Após a execução sistemática de quatro paradigmas distintos de modelação e a subsequente avaliação rigorosa da sua performance através de métricas discriminativas e de classificação, torna-se imperativo sintetizar os insights descobertos e formular recomendações estratégicas fundamentadas. Os resultados demonstraram que todos os modelos implementados excedem significativamente os critérios mínimos estabelecidos ($AUC > 0,8$ e $CA > 0,7$), validando a adequação da estratégia metodológica adotada e confirmando a viabilidade técnica de um sistema automatizado de identificação de oportunidades de *cross-selling*. Esta convergência de performance,

aliada às diferenças subtis, mas relevantes nos padrões de classificação observados nas matrizes de confusão, proporciona uma base sólida para a formulação de conclusões que transcendem a mera comparação quantitativa, abrangendo considerações de interpretabilidade, estabilidade de generalização e adequação ao contexto específico de negócio. As conclusões que se seguem integram não apenas a análise comparativa de performance, mas também a extração de conhecimento interpretável através das regras de decisão descobertas, culminando numa recomendação fundamentada para implementação em ambiente de produção que equilibra eficácia preditiva, robustez operacional e valor estratégico para a organização.

Na análise comparativa das métricas CA e AUC, emergem dois modelos com performance superior, um é o *Tree*, posicionado em primeiro lugar no *ranking*, e o *Logistic Regression*, ocupando a segunda posição, sendo que o modelo *Tree* apresenta sinais de ligeiro *overfitting*. Para contextos que privilegiem máxima confiabilidade preditiva, o *Tree* constitui a escolha superior, destacando-se pela menor degradação de CA, melhor performance final e uma degradação aceitável da AUC, garantindo um comportamento previsível em ambiente produtivo. Como alternativa versátil, o *Logistic Regression* oferece um equilíbrio robusto entre a AUC e CA, com degradação controlada em ambas as métricas, interpretabilidade superior e flexibilidade de ajuste que a tornam adequada para diversas aplicações.

3.6.2. DESCOBERTA DE PADRÕES: ALGORITMO *TREE*

O algoritmo *Tree* apresenta uma característica fundamental para este projeto de previsão comportamental em *cross-selling* de seguros de vida: a interpretabilidade algorítmica. Esta capacidade permite não apenas identificar clientes com maior propensão à aquisição, mas crucialmente compreender os fatores determinantes que influenciam a decisão de compra, fornecendo *insights* acionáveis para as equipas comerciais otimizarem as suas abordagens de venda e personalizarem as propostas de valor de acordo com os perfis específicos identificados.

A exploração sistemática das ramificações hierárquicas do modelo revela quais variáveis e combinações de características influenciam os cenários de propensão à conversão no contexto específico dos seguros de vida. Esta análise transcende a mera performance preditiva, transformando padrões comportamentais complexos em regras de negócio explícitas e diretamente acionáveis. Consequentemente, as equipas comerciais podem compreender não apenas "quem" apresenta maior probabilidade de conversão, mas crucialmente "porquê" determinados perfis demográficos, financeiros ou comportamentais demonstram maior recetividade aos produtos de seguro de vida, facilitando estratégias de segmentação mais eficazes e campanhas comerciais altamente direcionadas.

Não obstante os benefícios evidentes, é fundamental reconhecer as limitações inerentes ao algoritmo *Tree* que podem comprometer a robustez dos *insights* gerados. A principal fragilidade reside na instabilidade estrutural do modelo, podendo sofrer de *overfitting* em segmentos com amostras reduzidas e comprometer a generalização para novos contextos de mercado. Do ponto de vista

operacional, existe o risco de rigidez excessiva na aplicação das regras e dependência da qualidade dos dados históricos, que pode perpetuar enviesamentos existentes. Por conseguinte, a implementação desta abordagem exige validação contínua, monitorização regular da performance e flexibilidade estratégica para adaptar as regras extraídas à evolução dinâmica do contexto de mercado.

De seguida, apresentam-se os resultados mais relevantes da análise descritiva do algoritmo, destacando tanto os resultados com maior potencial, bem como aqueles que revelam pouco potencial de uso.

O resultado identificado como de elevada qualidade apresenta uma taxa de conversão de 92.8% (451/486 casos) e revela, através de uma sequência hierárquica dos três critérios que mais influenciam a variável *target* "*PossuiSeguroVida2024*". A arquitetura hierárquica estabelece-se da seguinte forma: **primeiro**, a condição primária é a ausência de apólice automóvel anterior a 2024 ("*PossuiApoliceAutoAntes2024: Não*"); **segundo**, o critério discriminante é a origem dos contactos ("*GrupoSourcesA: LeadsReferenciadas*"); **terceiro**, a condição final é o comportamento de comunicação escrita por parte do especialista ("*ComportamentoEspecialistaComunicacaoEscrita: EscritaNormal*"). Esta estrutura de árvore de decisão revela uma lógica de negócio clara e sequencial, onde a combinação específica destes três atributos cria um perfil de cliente com probabilidade de conversão superior a 90%. O modelo demonstra como variáveis aparentemente independentes se conjugam sinergicamente para formar um segmento de alta rentabilidade comercial, proporcionando insights valiosos para estratégias de *targeting* e otimização de campanhas (anexos: Imagem 5).

Um nó identificado com uma taxa de conversão de apenas 0.9% (6/654 casos), revelando, através de uma sequência hierárquica de dois critérios discriminantes, as variáveis que exercem influência negativa na variável *target* "*PossuiSeguroVida2024*". A arquitetura hierárquica estabelece-se da seguinte forma: **primeiro**, a condição primária é a existência de inscrições no CPCC ("*NumeroInscricoesCPCC > 0*"); **segundo**, o critério discriminante é o comportamento de comunicação escrita por parte do especialista ("*ComportamentoESpecialistaComunicacaoEscrita: PoucoEscrita*"). Este resultado oferece *insights* valiosos sobre os perfis de cliente com menor propensão à conversão, indicando claramente os padrões a evitar nas estratégias comerciais (anexos: Imagem 6).

Por último, apresento dois exemplos de resultados que evidenciam baixa qualidade. No primeiro caso, o algoritmo produz um resultado baseado em apenas 2 observações (2/0 casos), apresentando uma taxa de conversão aparentemente perfeita de 100%. No segundo exemplo, identifica-se um nó com apenas 3 casos (2/1 casos), resultando numa taxa de conversão de 66.7%. Ambos os resultados carecem de qualquer relevância estatística ou valor preditivo, uma vez que o reduzido número de casos em cada nó torna as conclusões estatisticamente insignificantes e inadequadas para fundamentar decisões estratégicas ou operacionais.

Esta arquitetura de árvore de decisão revela uma lógica de negócio clara e sequencial, proporcionando *insights* estratégicos valiosos para a otimização de recursos comerciais. Por um lado, permitiu identificar um resultado que delinea um perfil de cliente com probabilidade de conversão superior a 90%, permitindo a concentração de esforços comerciais nos segmentos mais promissores. Por outro lado, o algoritmo forneceu informação crucial sobre um perfil de cliente no qual não é recomendável investir recursos, uma vez que apresenta uma probabilidade de conversão substancialmente reduzidas, possibilitando assim uma alocação mais eficiente dos recursos disponíveis e uma estratégia comercial mais direcionada e rentável. Estas descobertas revelam uma oportunidade estratégica de *cross-selling* altamente direcionada, permitindo identificar com precisão os perfis de clientes que demonstram maior receptividade à aquisição de seguros de vida.

3.6.1. SELEÇÃO DO MODELO ÓTIMO

O *Logistic Regression* emerge como a solução equilibrada, oferecendo um compromisso ótimo entre performance e estabilidade de generalização com os segundos melhores resultados em ambas as métricas, e menor degradação treino-teste. A robustez comprovada da *Logistic Regression* no setor segurador, combinada com a sua interpretabilidade essencial para decisões estratégicas de *cross-selling* e possui menor propensão ao *overfitting*, em relação ao *Tree*, posicionando-o como a escolha mais prudente para implementação produtiva, onde a consistência, transparência algorítmica e capacidade de generalização são fatores críticos de sucesso empresarial.

Ao analisar os resultados da matriz de confusão observa-se um resultado notável: o *Naive Bayes* demonstra uma capacidade excepcional de identificar verdadeiros positivos (99.5%), conseguindo prever com elevada precisão os clientes com propensão para *cross-selling* de seguros de vida. Contudo, apresenta limitações na identificação de verdadeiros negativos, gerando um número considerável de falsos positivos. Apesar desta limitação, em contextos comerciais onde cada oportunidade perdida representa potencial receita não realizada, a elevada capacidade de deteção de clientes interessados posiciona o *Naive Bayes* como a solução ótima para estratégias que privilegiem a maximização de alcance e cobertura, mesmo aceitando um maior número de recomendações incorretas.

Considerando o contexto de negócio e os objetivos do projeto, o *Naive Bayes* emerge como a escolha principal. A elevada sensibilidade do *Naive Bayes* para identificar potenciais compradores constitui precisamente a vantagem estratégica procurada no contexto de *cross-selling*. A preferência por um modelo que maximize a captação de oportunidades comerciais, mesmo gerando alguns falsos positivos, alinha-se perfeitamente com os objetivos de expansão da carteira de seguros de vida. Paradoxalmente, mesmo os falsos positivos podem transformar-se numa vantagem estratégica de médio prazo, funcionando como um processo de maturação de conhecimento junto do cliente que aumenta a literacia financeira (Fisch & Seligman, 2022) e, conseqüentemente, eleva a probabilidade de conversão em contactos futuros (Weedige et al., 2019).

Todavia, o *Naive Bayes* não constitui a única solução viável. Os resultados obtidos pelo algoritmo *Tree*, permite analisar os nós de decisão e identificar perfis de clientes com maior e menor propensão à aquisição de seguros de vida, tornam-no um modelo complementar essencial para personalizar a experiência do cliente e desenvolver abordagens comerciais mais direcionadas e eficazes.

4. CONSIDERAÇÕES FINAIS e DISCUSSÃO

Este projeto estabelece um marco transformacional no *cross-selling* de seguros de vida, demonstrando como a aplicação estratégica de algoritmos de *machine learning* - especificamente modelos *Naive Bayes* e *Tree* com precisão superior a 90% - transforma significativamente a capacidade de identificar oportunidades comerciais personalizadas numa MSP portuguesa. A convergência entre dados comportamentais e técnicas preditivas não apenas colmata lacunas científicas críticas na literatura sobre personalização no setor segurador, mas materializa-se em impacto comercial mensurável: elevação significativa dos níveis de satisfação do cliente através de ofertas altamente personalizadas, potenciação do *engagement* mediante interações mais relevantes e oportunas, fortalecimento da retenção de clientes através de experiências diferenciadas, consolidação da fidelização pela criação de valor percebido superior, e incremento substancial das vendas através de estratégias de *cross-selling* cientificamente fundamentadas. Esta abordagem maximiza o *customer lifetime value*, otimiza radicalmente os custos de aquisição, e cria relacionamentos cliente-empresa duradouros que reduzem substancialmente o *churn*. Os resultados transcendem o âmbito académico, fornecendo metodologias cientificamente validadas que respondem diretamente às necessidades de transformação digital identificadas pela Autoridade de Supervisão de Seguros e Fundos de Pensões, (2024) e posicionam o setor segurador português na vanguarda da inovação europeia através da transformação de dados em vantagem competitiva sustentável.

A arquitetura de *matching* inteligente desenvolvida demonstra como as MSPs possuem vantagens informacionais únicas que transcendem os silos informacionais das seguradoras tradicionais, transformando dados transversais sobre comportamentos de compra em ativos estratégicos que sustentam uma vantagem competitiva duradoura (Pousttchi & Gleiss, 2019; Teece et al., 2022). A utilização de múltiplas dimensões de dados - comportamentais, transacionais e demográficos - permite criar experiências altamente contextualizadas que respondem às circunstâncias específicas de cada cliente, enquanto a automação do processo de identificação de oportunidades elimina as fricções tradicionais no *cross-selling*, reduzindo os custos de transação e maximizando os benefícios das plataformas multi-sided (Chan et al., 2022; Micallef et al., 2023; Verhoef et al., 2021). Simultaneamente, cada interação gera dados que alimentam o modelo, criando um ciclo virtuoso de melhoria contínua que fortalece a posição competitiva da plataforma (Suuronen et al., 2024).

A transformação operacional do negócio materializa-se através da capacidade de identificar e conectar clientes com produtos adequados no momento certo ao longo da vida do cliente. A sensibilidade

excepcional do modelo *Naive Bayes*, conjugada com os perfis de cliente identificados no modelo *Tree*, permite à Doutor Finanças implementar uma estratégia de *cross-selling* de seguros de vida que maximiza o *customer lifetime value* através da introdução de produtos de maior valor agregado (Winer, 2001), enquanto otimiza radicalmente a estrutura de custos comerciais pela eliminação dos custos de captação de novos clientes (De Caigny et al., 2018; Zhang et al., 2022). Os resultados obtidos transcendem a personalização reativa tradicional, evoluindo para uma abordagem preditiva que antecipa necessidades antes mesmo do cliente as expressar conscientemente, criando uma experiência proativa que supera as expectativas tradicionais (Ponsignon, 2023). Esta transformação operacional concentra os investimentos comerciais exclusivamente nos custos de conversão, que são significativamente inferiores, melhorando substancialmente as margens operacionais e o retorno sobre investimento.

A revolução da experiência do cliente concretiza-se através de comunicações altamente personalizadas e contextualizadas que impactam cada cliente no momento preciso, substituindo abordagens genéricas por interações que respondem às necessidades específicas e ao timing ideal de cada perfil (McLeod, 2025). A capacidade de retreino automático assegura personalização dinâmica que evolui com as mudanças comportamentais dos clientes, permitindo alterar a abordagem ao mesmo cliente ao longo da sua vida para proporcionar a melhor experiência possível. Esta personalização permite realizar *cross-selling* sem intrusividade, constituindo uma *mais-valia* que chega ao cliente no timing ideal e mantém a relação de confiança (Li et al., 2011). Mesmo nos casos de falsos positivos, a estratégia transforma-se numa oportunidade de educação financeira (Weedige et al., 2019), partilhando conhecimento sobre a importância dos seguros de vida no bem-estar financeiro das famílias, criando um processo de maturação que pode resultar em conversões futuras.

O impacto estratégico materializa-se na transformação profunda do relacionamento cliente-plataforma, onde o *cross-selling* de seguros de vida através de uma abordagem personalizada revoluciona a dinâmica relacional, elevando os níveis de confiança através de interações relevantes e contextualizadas que demonstram genuína preocupação com o bem-estar financeiro do cliente (Bhatia et al., 2024; Fisch & Seligman, 2022; C. Lin et al., 2017; Weedige et al., 2019). Esta personalização emerge não como objetivo isolado, mas como consequência natural da aplicação inteligente de IA aos dados comportamentais e transacionais, estabelecendo um novo standard de excelência na experiência cliente que apenas é possível através da convergência tecnológica alcançada neste projeto. Esta evolução fortalece os laços relacionais e reduz as taxas de *churn*, enquanto o aumento dos *touchpoints* gera um fluxo contínuo de dados comportamentais que alimenta constantemente os algoritmos preditivos, criando um ciclo virtuoso de melhoria contínua da precisão das previsões e otimização progressiva da experiência (Nagaraju et al., 2023; Ortega et al., 2023; Qaraeen et al., 2024).

A culminação desta transformação reflete-se numa melhoria da performance comercial global que impacta diretamente o crescimento através do aumento das comissões, melhoria dos rácios de conversão, e criação de um modelo sustentável baseado na maximização do valor da base de clientes existente. Os

resultados demonstram o potencial transformacional desta abordagem na melhoria do nível de satisfação do cliente através da oferta de produtos verdadeiramente alinhados com as suas necessidades, no incremento do *engagement* através de interações mais relevantes e personalizadas (Brodie et al., 2011; Hollebeek, 2011), na otimização da retenção e fidelização mediante a criação de experiências diferenciadas que fortalecem o relacionamento cliente-plataforma, e na maximização do retorno comercial através da identificação precisa de oportunidades de *cross-selling* com elevada probabilidade de conversão. A capacidade de antecipar necessidades e conectar clientes com soluções adequadas no momento certo estabelece um ciclo virtuoso (Li et al., 2011) onde a personalização se traduz simultaneamente em maior valor percebido pelo cliente e em performance comercial superior para a organização, permitindo que as diferentes áreas organizacionais revolucionem a experiência através da personalização: desde a criação de campanhas e *landing pages* com conteúdo direcionado para cada perfil, até à abordagem comercial personalizada, estabelecendo um novo standard de excelência que transcende o âmbito académico e se materializa em vantagem competitiva sustentável no mercado segurador português.

5. RECOMENDAÇÕES PARA INVESTIGAÇÕES FUTURAS

Com base nos resultados obtidos, recomendam-se desenvolvimentos investigacionais específicos que aprofundem e expandam o conhecimento gerado por este estudo. Primeiramente, sugere-se a realização de um novo estudo com uma base de dados substancialmente superior para validar se o modelo *Naive Bayes* e *Tree* mantém consistentemente o nível de qualidade de desempenho na identificação de verdadeiros positivos, permitindo confirmar a robustez e escalabilidade dos resultados obtidos. Simultaneamente, recomenda-se a execução de testes mais extensivos ao modelo *kNN* com diferentes valores de "*k*" *neighbors*, procurando identificar configurações que reduzam efetivamente o *overfitting* e otimizem o desempenho preditivo. O alargamento do estudo a outros algoritmos avançados, nomeadamente *gradient boosting* e *neural networks*, constituirá uma contribuição valiosa para estabelecer *benchmarks* comparativos mais abrangentes e identificar potenciais melhorias na precisão preditiva. Na dimensão prática, torna-se fundamental estudar os resultados pós-implementação para avaliar a aceitação organizacional ao modelo, analisando capacidade de adaptação dos processos internos, a integração com os sistemas existentes, e a evolução progressiva da implementação ao longo do tempo, identificando potenciais resistências, barreiras operacionais e oportunidades de otimização que emergem durante a transição para um paradigma de vendas orientado por inteligência artificial. Finalmente, recomenda-se uma análise longitudinal dos resultados após implementação para medir empiricamente a reação dos clientes e verificar se ocorreu efetivamente o aumento dos níveis esperados de satisfação, *engagement*, retenção, fidelização e retorno comercial, através de métricas quantitativas e qualitativas que permitam avaliar o impacto real da personalização preditiva na experiência do cliente e na performance comercial da organização, validando assim o impacto transformacional previsto e

fornecendo insights estratégicos para otimizações futuras do modelo preditivo e refinamento das estratégias de implementação.

6. LIMITAÇÕES

Esta investigação reconhece várias limitações metodológicas e técnicas que podem ter influenciado os resultados obtidos e que devem ser consideradas na interpretação dos resultados.

A qualidade dos dados constituiu uma limitação significativa, com *missing data* situando-se entre 23% e 25% na maioria das variáveis, sendo que a variável "*distritos*" apresentou valores em falta superiores a 30%. Esta elevada percentagem de dados ausentes pode ter gerado enviesamento nos resultados do modelo, comprometendo a representatividade e a precisão das previsões geradas.

A dimensão da base de dados utilizada representa outra limitação relevante, uma vez que uma base de dados superior proporcionaria maior robustez aos resultados e permitiria validações mais consistentes dos padrões identificados pelos algoritmos de *machine learning*. Esta limitação é particularmente relevante considerando a complexidade dos comportamentos de consumo no setor segurador e a necessidade de capturar variações sazonais e tendências de longo prazo.

Do ponto de vista técnico, não foi possível incluir as variáveis *Net Promoter Score (NPS)* e *Customer Satisfaction Score (CSAT)* na base de dados devido a limitações técnicas. Esta ausência constitui uma lacuna significativa, uma vez que estas métricas forneceriam informação valiosa sobre o *engagement* e satisfação do cliente.

A visualização dos nós do modelo *Tree* apresentou limitações na identificação de perfis diferenciados para cross-selling. Esta limitação decorre do equilíbrio entre complexidade do modelo e capacidade de generalização. *Datasets* de maior dimensão poderiam revelar segmentos adicionais com volume suficiente para aplicação comercial.

Durante o processo de preparação dos dados, identificaram-se inconsistências que exigiram correções manuais, nomeadamente situações onde clientes eram indicados como não tendo apólices ativas antes de 2024, mas simultaneamente apresentavam registos de apólices ativas nesse período. Estas inconsistências podem indicar problemas na qualidade dos dados originais que, apesar das correções efetuadas, podem ter introduzido ruído nos modelos.

Finalmente, os problemas de *overfitting* constituíram uma limitação crítica, com particular relevância nos resultados do modelo *kNN*. Esta situação indica que os modelos podem ter memorizado padrões específicos dos dados de treino, resultando numa perda significativa da qualidade de previsão quando aplicados a dados reais. A resolução desta limitação requer novas iterações no modelo *kNN*, uma base de dados mais extensa e a inclusão de mais variáveis independentes, elementos que contribuirão para reduzir o *overfitting* e melhorar a capacidade de generalização dos algoritmos desenvolvidos.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- Aka, D. O., Kehinde, O. J., & Ogunnaike, O. O. (2016). Relationship Marketing and Customer Satisfaction: A Conceptual Perspective. *Binus Business Review*, 7(2), 185–190. <https://doi.org/10.21512/bbr.v7i2.1502>
- Akhavan, F., & Hassannayebi, E. (2024). A hybrid machine learning with process analytics for predicting customer experience in online insurance services industry. *Decision Analytics Journal*, 11. <https://doi.org/10.1016/j.dajour.2024.100452>
- Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2024). AI-Powered Innovation in Digital Transformation: Key Pillars and Industry Impact. *Sustainability (Switzerland)*, 16(5), 1–25. <https://doi.org/10.3390/su16051790>
- Alexander, F. J., Lin, M., Qian, X., & Yoon, B. J. (2023). Accelerating scientific discoveries through data-driven innovations. *Patterns*, 4(11), 100876. <https://doi.org/10.1016/j.patter.2023.100876>
- Alijoyo, F. A., Aziz, T. S. A., Omer, N., Yusuf, N., Kumar, M. D., Ramesh, A. V. N., Ulmas, Z., & Baker El-Ebiary, Y. A. (2025). Personalized marketing: Leveraging AI for culturally aware segmentation and targeting. *Alexandria Engineering Journal*, 119, 8–21. <https://doi.org/10.1016/j.aej.2025.01.074>
- Allweins, M. M., Proesch, M., & Ladd, T. (2021). The Platform Canvas—Conceptualization of a Design Framework for Multi-Sided Platform Businesses. *Entrepreneurship Education and Pedagogy*, 4(3), 455–477. <https://doi.org/10.1177/2515127420959051>
- Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: Personalization & customization of services. In *Applied Computing and Informatics* (Vol. 15, Issue 2, pp. 94–101). Elsevier B.V. <https://doi.org/10.1016/j.aci.2018.05.004>
- Aravind, C., Sivanesan, S., & Singh, R. (2023). Reinforced learning experience framework. In *AIP Conference Proceedings* (Vol. 2643). <https://doi.org/10.1063/5.0114287>
- Autoridade de Supervisão de Seguros e Fundos de Pensões. (2024). *RELATÓRIO DO SETOR SEGURADOR E DOS FUNDOS DE PENSÕES 2023*. <https://www.asf.com.pt>
- Babatunde, S. O., Odejide, O. A., Edunjobi, T. E., & Ogundipe, D. O. (2024). THE ROLE OF AI IN MARKETING PERSONALIZATION: A THEORETICAL EXPLORATION OF CONSUMER ENGAGEMENT STRATEGIES. *International Journal of Management & Entrepreneurship Research*, 6(3), 936–949. <https://doi.org/10.51594/ijmer.v6i3.964>
- Becker, W. E., Shapiro, C., & Varian, H. R. (1999). Information Rules: A Strategic Guide to the Network Economy. *The Journal of Economic Education*, 30(2), 189. <https://doi.org/10.2307/1183273>

- Belhadi, A., Abdellah, N., & Nezai, A. (2023). The Effect of Big Data on the Development of the Insurance Industry. *Business Ethics and Leadership*, 7(1), 1–11. [https://doi.org/10.21272/bel.7\(1\).1-11.2023](https://doi.org/10.21272/bel.7(1).1-11.2023)
- Berry, L. (2002). Relationship Marketing of Services Perspectives from 1983 and 2000. *Journal of Relationship Marketing*, 1, 59–77. https://doi.org/10.1300/J366v01n01_05
- Berry, L. L. (1995). Relationship Marketing of Services-Growing Interest, Emerging Perspectives. *Journal of the Academy of Marketing Science*, 23(4), 236–245. <https://doi.org/doi.org/10.1177/009207039502300402>
- Bhatia, R., Bhat, A. K., & Tikoria, J. (2021). Life insurance purchase behaviour: A systematic review and directions for future research. In *International Journal of Consumer Studies* (Vol. 45, Issue 6, pp. 1149–1175). John Wiley and Sons Inc. <https://doi.org/10.1111/ijcs.12681>
- Bhatia, R., Bhat, A. K., & Tikoria, J. (2024). Empowering Informed Life Insurance Decisions: The Impact of Financial Literacy on Framing Effects. *Services Marketing Quarterly*, 45(4), 429–457. <https://doi.org/10.1080/15332969.2024.2415759>
- Bhattacharya, C. B., & Sen, S. (2003). Consumer-Company Identification: A Framework for Understanding Consumers' Relationships with Companies. In *Journal of Marketing* (Vol. 67).
- Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision trees: from efficient prediction to responsible AI. In *Frontiers in Artificial Intelligence* (Vol. 6). Frontiers Media SA. <https://doi.org/10.3389/frai.2023.1124553>
- Bolton, R., & Lemon, K. (1999). A Dynamic Model of Customers' Usage of Services: Usage as an Antecedent and Consequence of Satisfaction. *Journal of Marketing Research*, 36, 171–186. <https://doi.org/10.1177/002224379903600203>
- Borden, N. H. (1964). The Concept of the Marketing Mix. *Journal of Advertising Research*, 2, 7–12.
- Boudet, J., Gregg, B., Rathje, K., Stein, E., & Vollhardt, K. (2019, June). The future of personalization- and how to get ready for it. *McKinsey & Company*. <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-future-of-personalization-and-how-to-get-ready-for-it>
- Boyatzis, R. E. , & G. D. , & M. A. (2002). *Primal Leadership: Realizing the Power of Emotional Intelligence*. Harvard Business Review Press. <https://www.researchgate.net/publication/230854764>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

- Brodie, R. J., Hollebeek, L. D., Jurić, B., & Ilić, A. (2011). Customer engagement: Conceptual domain, fundamental propositions, and implications for research. *Journal of Service Research*, 14(3), 252–271. <https://doi.org/10.1177/1094670511411703>
- Catlin, T., Lorenz, J.-T., Nandan, J., Sharma, S., & Waschto, A. (2018, January). Insurance beyond digital: The rise of ecosystems and platforms. *McKinsey & Company*, 1–13. <https://www.the-digital-insurer.com/>
- Chan, H., Yang, M. X., & Zeng, K. J. (2022). Bolstering ratings and reviews systems on multi-sided platforms: A co-creation perspective. *Journal of Business Research*, 139, 208–217. <https://doi.org/10.1016/j.jbusres.2021.09.052>
- Chandra, S., Verma, S., Lim, W. M., Kumar, S., & Donthu, N. (2022). Personalization in personalized marketing: Trends and ways forward. In *Psychology and Marketing* (Vol. 39, Issue 8, pp. 1529–1562). John Wiley and Sons Inc. <https://doi.org/10.1002/mar.21670>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). *CRISP-DM 1.0 Step-by-step data mining guide*. DaimlerChrysler.
- Chen, Y., & Prentice, C. (2025). Integrating Artificial Intelligence and Customer Experience. *Australasian Marketing Journal*, 33(2), 141–153. <https://doi.org/10.1177/14413582241252904>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1). <https://doi.org/10.1186/s12864-019-6413-7>
- Christensen, C. M., Hall, T., Dillon, K., & Duncan, D. S. (2016). *Competing Against Luck: The Story of Innovation and Customer Choice*. HarperCollins. https://books.google.pt/books?id=zGd_CwAAQBAJ
- Davidson, H., & O'Reilly, T. (1987). *Offensive Marketing, Or, How to Make Your Competitors Followers*. Penguin Books. <https://books.google.pt/books?id=9XzKAAAACAAJ>
- De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 760–772. <https://doi.org/10.1016/j.ejor.2018.02.009>
- Domingos, P., & Pazzani, M. (1997). On the Optimality of the Simple Bayesian Classifier under Zero-OneLoss. *Machine Learning*, 29, 103–130.
- Dong, X., & McIntyre, S. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. *Quantitative Finance*, 14. <https://doi.org/10.1080/14697688.2014.946440>

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37–54. <https://doi.org/10.1609/aimag.v17i3.1230>
- Fisch, J. E., & Seligman, J. S. (2022). Trust, financial literacy, and financial market participation. *Journal of Pension Economics and Finance*, 21(4), 634–664. <https://doi.org/10.1017/S1474747221000226>
- Fournier, S. (1998). Consumers and Their Brands: Developing Relationship Theory in Consumer Research. *Journal of Consumer Research*, 24(4), 343–353. <https://doi.org/10.1086/209515>
- Gahler, M., Klein, J. F., & Paul, M. (2023). Customer Experience: Conceptualization, Measurement, and Application in Omnichannel Environments. *Journal of Service Research*, 26(2), 191–211. <https://doi.org/10.1177/10946705221126590>
- Gandhudi, M., Alphonse, P. J. A., Velayudham, V., Nagineni, L., & Gangadharan, G. (2025). Dynamic customer behavior prediction in subscription services using causal reinforcement learning. *Engineering Applications of Artificial Intelligence*, 155. <https://doi.org/10.1016/j.engappai.2025.111030>
- Ganti, V., & Sarma, A. (2013). Data Cleaning: A Practical Perspective. *Synthesis Lectures on Data Management*, 5, 1–85. <https://doi.org/10.2200/S00523ED1V01Y201307DTM036>
- Garrido-Moreno, A., & Padilla-Meléndez, A. (2011). Analyzing the impact of knowledge management on CRM success: The mediating effects of organizational factors. *International Journal of Information Management*, 31(5), 437–444. <https://doi.org/10.1016/j.ijinfomgt.2011.01.002>
- Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, 4, 1–58. <https://doi.org/10.1162/neco.1992.4.1.1>
- Ghoshal, A., Mookerjee, V. S., & Sarkar, S. (2021). Recommendations and Cross-selling: Pricing Strategies when Personalizing Firms Cross-sell. *Journal of Management Information Systems*, 38(2), 430–456. <https://doi.org/10.1080/07421222.2021.1912930>
- Gupta, S., Pansari, A., & Kumar, V. (2018). Global Customer Engagement. *Journal of International Marketing*, 26(1), 4–29. <https://doi.org/10.1509/jim.17.0091>
- Hanelt, A., Bohnsack, R., Marz, D., & Antunes Marante, C. (2021). A Systematic Review of the Literature on Digital Transformation: Insights and Implications for Strategy and Organizational Change. *Journal of Management Studies*, 58(5), 1159–1197. <https://doi.org/10.1111/joms.12639>
- Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2004). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Math. Intell.*, 27, 83–85. <https://doi.org/10.1007/BF02985802>
- Hermann, E., & Puntoni, S. (2024). Artificial intelligence and consumer behavior: From predictive to generative AI. *Journal of Business Research*, 180. <https://doi.org/10.1016/j.jbusres.2024.114720>

- Hollebeek, L. D. (2011). Demystifying customer brand engagement: Exploring the loyalty nexus. In *Journal of Marketing Management* (Vol. 27, Issues 7–8, pp. 785–807). <https://doi.org/10.1080/0267257X.2010.500132>
- Hollebeek, L. D., Conduit, J., & Brodie, R. J. (2016). Strategic drivers, anticipated and unanticipated outcomes of customer engagement. In *Journal of Marketing Management* (Vol. 32, Issues 5–6, pp. 393–398). Routledge. <https://doi.org/10.1080/0267257X.2016.1144360>
- Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014). Consumer brand engagement in social media: Conceptualization, scale development and validation. *Journal of Interactive Marketing*, 28(2), 149–165. <https://doi.org/10.1016/j.intmar.2013.12.002>
- Hongladarom, S. (2023). Shoshana Zuboff, The age of surveillance capitalism: the fight for a human future at the new frontier of power. *AI & SOCIETY*, 38(6), 2359–2361. <https://doi.org/10.1007/s00146-020-01100-0>
- Hu, L., Chen, J., Vaughan, J., Yang, H., Wang, K., Sudjianto, A., & Nair, V. N. (2020). *Supervised machine learning techniques: An overview with applications to banking corporate model risk [Technical report]. Wells Fargo.*
- Huang, M.-H., Rust, R., & Maksimovic, V. (2019). The Feeling Economy: Managing in the Next Generation of Artificial Intelligence (AI). *California Management Review*, 61, 000812561986343. <https://doi.org/10.1177/0008125619863436>
- Ijomah, T. I., Idemudia, C., Eyo-Udo, N. L., & Anjorin, K. F. (2024). The role of big data analytics in customer relationship management: Strategies for improving customer engagement and retention. *World Journal of Advanced Science and Technology*, 6(1), 013–024. <https://doi.org/10.53346/wjast.2024.6.1.0038>
- Ka, V., Mohanasundaram, K., & Pothiyachi, V. (2023). *Regression tasks for machine learning* (pp. 133–157). <https://doi.org/10.1016/B978-0-323-91776-6.00009-9>
- Kanungo, R., Liu, R., & Gupta, S. (2024). Cognitive analytics enabled responsible artificial intelligence for business model innovation: A multilayer perceptron neural networks estimation. *Journal of Business Research*, 182. <https://doi.org/10.1016/j.jbusres.2024.114788>
- Kapoor, S., Wills, A. G., Hendriks, J., & Blackhall, L. (2025). Estimation of distribution grid line parameters using smart meter data with missing measurements. *International Journal of Electrical Power and Energy Systems*, 168. <https://doi.org/10.1016/j.ijepes.2025.110634>
- Karthick, V., Prabhakaran, J., Banu, P., Senthil Kumar, U. S., Vadivel, K., & Banu, Er. P. (2023). *Systematic Literature Review on GRC-A Study on Best Practices and Implementation Strategy in GRC*. 16(4), 2581–3986.

- Kasabov, E. (2015). *Marketing Mix*. <https://doi.org/10.1002/9781118785317.weom090307>
- Katz, E., Blumler, J. A. Y. G., & Gurevitch, M. (1973). USES AND GRATIFICATIONS RESEARCH. *Public Opinion Quarterly*, 37(4), 509–523. <https://doi.org/10.1086/268109>
- Katz, M. L., & Shapiro, C. (1985). Network Externalities, Competition, and Compatibility. *The American Economic Review*, 75(3), 424–440.
- Kitsara, I. (2022). Artificial Intelligence and the Digital Divide: From an Innovation Perspective. In A. Bounfour (Ed.), *Platforms and Artificial Intelligence : The Next Generation of Competences* (pp. 245–265). Springer International Publishing. https://doi.org/10.1007/978-3-030-90192-9_12
- Kozlenkova, I. V., Samaha, S. A., & Palmatier, R. W. (2014). Resource-based theory in marketing. In *Journal of the Academy of Marketing Science* (Vol. 42, Issue 1, pp. 1–21). Springer Science and Business Media, LLC. <https://doi.org/10.1007/s11747-013-0336-7>
- Kumar, A., Bhushan, S., Pokhrel, R., Al-Omari, A. I., Alanzi, A. R. A., & Alshqaq, S. S. (2025). Imputation of missing data for domain mean estimation using simple random sampling. *Kuwait Journal of Science*, 52(4). <https://doi.org/10.1016/j.kjs.2025.100461>
- Kumar, V., Rajan, B., Venkatesan, R., & Lecinski, J. (2019). Understanding the role of artificial intelligence in personalized engagement marketing. *California Management Review*, 61(4), 135–155. <https://doi.org/10.1177/0008125619859317>
- Kumar, V., & Reinartz, W. (2016). Creating enduring customer value. *Journal of Marketing*, 80(6), 36–68. <https://doi.org/10.1509/jm.15.0414>
- Kwak, H., Anderson, R. E., Leigh, T. W., & Bonifield, S. D. (2019). Impact of salesperson macro-adaptive selling strategy on job performance and satisfaction. *Journal of Business Research*, 94, 42–55. <https://doi.org/10.1016/j.jbusres.2018.09.015>
- Lambillotte, L., Bart, Y., & Poncin, I. (2022). When Does Information Transparency Reduce Downside of Personalization? Role of Need for Cognition and Perceived Control. *Journal of Interactive Marketing*, 57(3), 393–420. <https://doi.org/10.1177/10949968221095557>
- Ledro, C., Nosella, A., Vinelli, A., Dalla Pozza, I., & Souverain, T. (2025). Artificial intelligence in customer relationship management: A systematic framework for a successful integration. *Journal of Business Research*, 199. <https://doi.org/10.1016/j.jbusres.2025.115531>
- Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6), 69–96. <https://doi.org/10.1509/jm.15.0420>
- Leung, F. F., Gu, F. F., Li, Y., Zhang, J. Z., & Palmatier, R. W. (2022). Influencer Marketing Effectiveness. *Journal of Marketing*, 86(6), 93–115. <https://doi.org/10.1177/00222429221102889>

- Levitt, T. (1960, July). Harvard Business Review MARKETING MYOPIA. *Harvard Business Review*, 45–56. <https://hbr.org/2004/07/marketing-myopia>
- Levitt, T. (1980, January). Marketing Success Through Differentiation-of Anything Harvard Business Review. *Harvard Business Review*, 1–9. <https://hbr.org/1980/01/marketing-success-through-differentiation-of-anything>
- Li, S., Sun, B., & Montgomery, A. L. (2011). li-et-al-2011-cross-selling-the-right-product-to-the-right-customer-at-the-right-time. *Journal of Marketing Research*, Vol. XLVIII. <https://doi.org/10.1509/JMKR.48.4.683>
- Lim, S. T., Yuan, J. Y., Khaw, K. W., & Chew, X. (2023). Predicting Travel Insurance Purchases in an Insurance Firm through Machine Learning Methods after COVID-19. *Journal of Informatics and Web Engineering*, 2(2), 43–58. <https://doi.org/10.33093/jiwe.2023.2.2.4>
- Lin, C., Hsiao, Y. J., & Yeh, C. Y. (2017). Financial literacy, financial advisors, and information sources on demand for life insurance. *Pacific Basin Finance Journal*, 43, 218–237. <https://doi.org/10.1016/j.pacfin.2017.04.002>
- Lin, Y. T. J., Chang, C. Y., Cheng, S. Y., & Lin, M. Y. T. (2023). Probabilistic Customer Purchase Evolution Graph. *IEEE Access*, 11, 32962–32971. <https://doi.org/10.1109/ACCESS.2023.3263729>
- Linoff, G. S., & Berry, M. J. A. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (3rd ed.). Wiley Publishing, Inc.
- Luo, S. T., Cheng, B. W., & Hsieh, C. H. (2009). Prediction model building with clustering-launched classification and support vector machines in credit scoring. *Expert Systems with Applications*, 36(4), 7562–7566. <https://doi.org/10.1016/j.eswa.2008.09.028>
- Mahadevkar, S. V., Khemani, B., Patil, S., Kotecha, K., Vora, D. R., Abraham, A., & Gabralla, L. A. (2022). A Review on Machine Learning Styles in Computer Vision - Techniques and Future Directions. In *IEEE Access* (Vol. 10, pp. 107293–107329). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2022.3209825>
- Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. In *Knowledge Engineering Review* (Vol. 25, Issue 2, pp. 137–166). <https://doi.org/10.1017/S0269888910000032>
- McFall, L., Meyers, G., & Hoyweghen, I. Van. (2020). Editorial: The personalisation of insurance: Data, behaviour and innovation. In *Big Data and Society* (Vol. 7, Issue 2). SAGE Publications Ltd. <https://doi.org/10.1177/2053951720973707>

- McLeod, S. (2025). *Albert Bandura's Social Learning Theory* [simplypsychology.org/bandura.html](https://www.simplypsychology.org/bandura.html).
<https://www.simplypsychology.org/bandura.html>
- Méndez-Aparicio, M. D., Jiménez-Zarco, A., Izquierdo-Yusta, A., & Blazquez-Resino, J. J. (2020). Customer Experience and Satisfaction in Private Insurance Web Areas. *Frontiers in Psychology, 11*. <https://doi.org/10.3389/fpsyg.2020.581659>
- Metcalf, R. M. (1995). Metcalf's Law: A network becomes more valuable as it reach. *InfoWorld, 17*(40), 53. <https://www.proquest.com/trade-journals/metcalfes-law-network-becomes-more-valuable-as/docview/194306234/se-2?accountid=38384>
- Micallef, M., Keränen, J., & Kokshagina, O. (2023). The (un)intended consequences of multi-sided platform adoption for different actors in business networks. *Industrial Marketing Management, 115*, 214–227. <https://doi.org/10.1016/j.indmarman.2023.09.018>
- Mlčúchová, M. (2022). A Review of Platform Business Models. In *Scientific Papers of the University of Pardubice, Series D: Faculty of Economics and Administration* (Vol. 30, Issue 1). University of Pardubice. <https://doi.org/10.46585/sp30011454>
- M.N., S., & Hossin, M. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process, 5*(2), 01–11. <https://doi.org/10.5121/ijdkp.2015.5201>
- Mukhopadhyay, S., Singh, R. K., & Jain, T. (2024). Developing big data enabled Marketing 4.0 framework. *International Journal of Information Management Data Insights, 4*(1). <https://doi.org/10.1016/j.jjime.2024.100214>
- Nagaraju, J., Sathwik, A. S., Saiteja, B., Challa, N. P., & Naseeba, B. (2023). Predicting Customer Churn in Insurance Industry Using Big Data and Machine Learning. *2023 1st International Conference on Advances in Electrical, Electronics and Computational Intelligence, ICAEECI 2023*. <https://doi.org/10.1109/ICAEECI58247.2023.10370876>
- Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems, 50*(3), 559–569. <https://doi.org/10.1016/j.dss.2010.08.006>
- Obi, J. C. (2023). A comparative study of several classification metrics and their performances on data. *World Journal of Advanced Engineering Technology and Sciences, 8*(1), 308–314. <https://doi.org/10.30574/wjaets.2023.8.1.0054>
- Oliver, R. (2010). Customer Satisfaction. In *Metal Finishing* (Vol. 100). <https://doi.org/10.1002/9781444316568.wiem03008>

- Olufunke Anne Alabi, Nnenna Ijeoma Okeke, Abbey Ngochindo Igwe, Onyeka Chrisanctus Ofodile, & Chikezie Paul-Mikki Ewim. (2024). Omni-channel customer experience framework: enhancing service delivery in SMEs. *World Journal of Advanced Research and Reviews*, 24(2), 655–670. <https://doi.org/10.30574/wjarr.2024.24.2.3335>
- Oluwaseyitan, R. C., Hashim, H., Raja Yusof, R. N., & Kok, T. K. (2023). Family and Life Insurance Purchase Decision-Making: The Significance of Personal Values. *International Journal of Academic Research in Business and Social Sciences*, 13(2). <https://doi.org/10.6007/ijarbss/v13-i2/16412>
- Orenga-Roglá, S., & Chalmeta, R. (2016). Social customer relationship management: taking advantage of Web 2.0 and Big Data technologies. *SpringerPlus*, 5(1). <https://doi.org/10.1186/s40064-016-3128-y>
- Ortega, M., Quintanilla, J., Ong, E. R., Ramos, R. M., & Trinidad, C. J. (2023). Asfalís: A Web-based System for Customer Retention Strategies Optimization of a Car Insurance Company using Cohort and Churn Analysis. *6th International Conference on Inventive Computation Technologies, ICICT 2023 - Proceedings*, 1065–1072. <https://doi.org/10.1109/ICICT57646.2023.10134149>
- Palmatier, R. W., Dant, R. P., Grewal, D., & Evans, K. R. (2006). Factors Influencing the Effectiveness of Relationship Marketing: A Meta-Analysis. *Journal of Marketing*, 70, 136–153.
- Palmatier, R. W., Houston, M. B., & Hulland, J. (2018). Review articles: purpose, process, and structure. In *Journal of the Academy of Marketing Science* (Vol. 46, Issue 1). Springer New York LLC. <https://doi.org/10.1007/s11747-017-0563-4>
- Papazafeiropoulou, A., & Spanaki, K. (2016). Understanding governance, risk and compliance information systems (GRC IS): The experts view. *Information Systems Frontiers*, 18(6), 1251–1263. <https://doi.org/10.1007/s10796-015-9572-3>
- Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform Revolution: How Networked Markets Are Transforming the Economy and How to Make Them Work for You*. W. W. Norton. <https://books.google.pt/books?id=Bvd1CQAAQBAJ>
- Patterson, P., Yu, T., & Kimpakorn, N. (2014). Killing two birds with one stone: Cross-selling during service delivery. *Journal of Business Research*, 67(9), 1944–1952. <https://doi.org/10.1016/j.jbusres.2013.11.013>
- Payne, A., & Frow, P. (2006). Customer Relationship Management: from Strategy to Implementation. *Journal of Marketing Management*, 22(1–2), 135–168. <https://doi.org/10.1362/026725706776022272>

- Peppers, D., & Rogers, M. (2016). Managing Customer Relationships: A Strategic Framework: Third Edition. In *Managing Customer Relationships: A Strategic Framework: Third Edition*. <https://doi.org/10.1002/9781119239833>
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19(C), 123–205. [https://doi.org/10.1016/S0065-2601\(08\)60214-2](https://doi.org/10.1016/S0065-2601(08)60214-2)
- Ponsignon, F. (2023). Making the customer experience journey more hedonic in a traditionally utilitarian service context: a case study. *Journal of Service Management*, 34(2), 294–315. <https://doi.org/10.1108/JOSM-03-2021-0096>
- Pousttchi, K., & Gleiss, A. (2019). Surrounded by middlemen - how multi-sided platforms change the insurance industry. *Electronic Markets*, 29(4), 609–629. <https://doi.org/10.1007/s12525-019-00363-w>
- Powers, D. M. W. (2011). EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION. *Journal of Machine Learning Technologies*, 2(1), 37–63. <https://doi.org/10.9735/2229-3981>
- Prasetyowati, S. S., & Sibaroni, Y. (2024). Unlocking the potential of Naive Bayes for spatio temporal classification: a novel approach to feature expansion. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-024-00958-x>
- Qaraeen, T., Qaqour, N., & Taqatqa, S. (2024). Predictive Customer Analytics: Machine Learning for Churn Prediction and Retention. *Ahliya Journal of Business Technology and MEAN Economies*, 1, 1–28.
- Qazi, M., Tollas, K., Kanchinadam, T., Bockhorst, J., & Fung, G. (2020). Designing and deploying insurance recommender systems using machine learning. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(4). <https://doi.org/10.1002/widm.1363>
- Quach, S., Thaichon, P., Martin, K. D., Weaven, S., & Palmatier, R. W. (2022). Digital technologies: tensions in privacy and data. *Journal of the Academy of Marketing Science*, 50(6), 1299–1323. <https://doi.org/10.1007/s11747-022-00845-y>
- Ramaj, A., & Ismaili, R. (2015). Customer Relationship Management, Customer Satisfaction and Loyalty. *Academic Journal of Interdisciplinary Studies*. <https://doi.org/10.5901/ajis.2015.v4n3s1p594>
- Rochet, J.-C., & Tirole, J. (2006). Two-Sided Markets: A Progress Report. *The RAND Journal of Economics*, 37(3), 645–667. <http://www.jstor.org/stable/25046265>

- Rohlf, J. H. (2003). *Bandwagon Effects in High-technology Industries*. MIT Press.
<https://books.google.pt/books?id=rmFag8P4CF8C>
- Rosário, A. T., & Casaca, J. A. (2023). Relational Marketing and Customer Satisfaction: A Systematic Literature Review. *Estudios Gerenciales*, 39(169), 516–532.
<https://doi.org/10.18046/j.estger.2023.167.6218>
- Roy, S. K., Tehrani, A. N., Pandit, A., Apostolidis, C., & Ray, S. (2025). Ai-capable relationship marketing: Shaping the future of customer relationships. *Journal of Business Research*, 192.
<https://doi.org/10.1016/j.jbusres.2025.115309>
- Rusnaini, S., M., A., Jessika, S., Pratiwi, W., Marlina, E., & Hamirul. (2024). The Customer Experience Revolution: Building Brand Loyalty in the Age of Digital Disruption. *Enigma in Economics*, 2(1), 69–80. <https://doi.org/10.61996/economy.v2i1.57>
- Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., & Edwards, D. D. (1995). *Artificial Intelligence A Modern Approach* (1st ed.). Prentice Hall.
- Sadiq, E., Alkarani, H., & Chakranarayan, V. (2025). Predictive Analytics in Customer Relationship Management. *2nd International Conference on IT Innovations and Knowledge Discovery, ITIKD 2024*. <https://doi.org/10.1109/ITIKD63574.2025.11005214>
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3), Article Number: e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Sanchez-Cartas, J. M., & León, G. (2021). MULTISIDED PLATFORMS AND MARKETS: A SURVEY OF THE THEORETICAL LITERATURE. *Journal of Economic Surveys*, 35(2), 452–487. <https://doi.org/10.1111/joes.12409>
- Shah, D., Kumar, V., Qu, Y., & Chen, S. (2012). Unprofitable Cross-Buying: Evidence from Consumer and Business Markets. *Journal of Marketing*, 76, 78–95.
<https://doi.org/doi.org/10.1509/jm.10.0445>
- Sheth, J. N., Jain, V., & Ambika, A. (2023). The growing importance of customer-centric support services for improving customer experience. *Journal of Business Research*, 164.
<https://doi.org/10.1016/j.jbusres.2023.113943>
- Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110. <https://doi.org/10.1016/j.ssresearch.2022.102817>
- Silva-Atencio, G. (2025). Data Analytics for Sales Strategies: Lessons in Customer Management. *Journal of Comprehensive Business Administration Research*.
<https://doi.org/10.47852/bonviewJCBAR52025199>

- Soofi, A. A., & Awan, A. (2017). Classification Techniques in Machine Learning: Applications and Issues. *Journal of Basic & Applied Sciences*, *13*, 459–465. <https://doi.org/10.6000/1927-5129.2017.13.76>
- Srivastava, A. K., & Jadid, A. (2020). An Empirical Study on Consumer Behavior of Life Insurance Purchasing Decision. *The Creative Launcher*, *5*(1), 16–31. <https://doi.org/10.53032/tcl.2020.5.1.04>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.
- Stourm, V., Neslin, S. A., Bradlow, E. T., Breugelmans, E., Chun, S. Y., Gardete, P., Kannan, P. K., Kopalle, P., Park, Y. H., Amariles, D. R., Thomadsen, R., Liu-Thompkins, Y., & Venkatesan, R. (2020). Refocusing loyalty programs in the era of big data: a societal lens paradigm. *Marketing Letters*, *31*(4), 405–418. <https://doi.org/10.1007/s11002-020-09523-x>
- Strömbäck, J. (2007). Political marketing and professionalized campaigning: A conceptual analysis. *Journal of Political Marketing*, *6*(2–3), 49–67. https://doi.org/10.1300/J199v06n02_04
- Suuronen, S., Ukko, J., Saunila, M., Rantala, T., & Rantanen, H. (2024). The implications of multi-sided platforms in managing digital business ecosystems. *Journal of Business Research*, *175*. <https://doi.org/10.1016/j.jbusres.2024.114544>
- Szabó, S., Holb, I. J., Abriha-Molnár, V. É., Szatmári, G., Singh, S. K., & Abriha, D. (2024). Classification Assessment Tool: A program to measure the uncertainty of classification models in terms of class-level metrics. *Applied Soft Computing*, *155*. <https://doi.org/10.1016/j.asoc.2024.111468>
- Teece, D. J., Pundziene, A., Heaton, S., & Vadi, M. (2022). Managing Multi-Sided Platforms: Platform Origins and Go-to-Market Strategy. *California Management Review*, *64*(4), 5–19. <https://doi.org/10.1177/00081256221109961>
- Tolga Akçura, M., & Srinivasan, K. (2005). Research note: Customer intimacy and cross-selling strategy. *Management Science*, *51*(6), 1007–1012. <https://doi.org/10.1287/mnsc.1050.0390>
- van Witteloostuijn, A., Vanderstraeten, J., Slabbinck, H., Dejardin, M., Hermans, J., & Coreynen, W. (2022). From explanation of the past to prediction of the future: A comparative and predictive research design in the Social Sciences. *Social Sciences and Humanities Open*, *6*(1). <https://doi.org/10.1016/j.ssaho.2022.100269>
- Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Qi Dong, J., Fabian, N., & Haenlein, M. (2021). Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research*, *122*, 889–901. <https://doi.org/10.1016/j.jbusres.2019.09.022>

- Verhoef, P. C., Lemon, K. N., Parasuraman, A., Roggeveen, A., Tsiros, M., & Schlesinger, L. A. (2009). Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of Retailing*, 85(1), 31–41. <https://doi.org/10.1016/j.jretai.2008.11.001>
- Wagner, A. (1891). Marshall's Principles of Economics. *The Quarterly Journal of Economics*, 5(3), 319–338. <https://doi.org/10.2307/1879612>
- Wang, H., & Raj, B. (2017). *On the Origin of Deep Learning*. <https://doi.org/10.48550/arXiv.1702.07800>
- Warrington, T. (2002). The Secrets of Word-of-Mouth Marketing: How to Trigger Exponential Sales through Runaway Word of Mouth. *Journal of Consumer Marketing*, 19(4), 364–366. <https://doi.org/10.1108/jcm.2002.19.4.364.4>
- Weedige, S. S., Ouyang, H., Gao, Y., & Liu, Y. (2019). Decision making in personal insurance: Impact of insurance literacy. *Sustainability (Switzerland)*, 11(23). <https://doi.org/10.3390/su11236795>
- Weidig, J., Weippert, M., & Kuehnl, C. (2024). Personalized touchpoints and customer experience: A conceptual synthesis. *Journal of Business Research*, 177. <https://doi.org/10.1016/j.jbusres.2024.114641>
- Winer, R. S. (2001). A Framework for Customer Relationship Management. *CALIFORNIA MANAGEMENT REVIEW*, 43.
- Xing, X., Huang, H., & Hedenstierna, C. P. T. (2023). Selling through online marketplaces with consumer profiling. *Journal of Business Research*, 164. <https://doi.org/10.1016/j.jbusres.2023.114022>
- Yang, Q. (2017). *The Role of Design in Creating Machine-Learning-Enhanced User Experience*.
- Zaki, M. J., & Meira Jr., W. (2014). *Data mining and analysis : fundamental concepts and algorithms* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511810114>
- Zeithaml, V. A. (1988). Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. *Journal of Marketing*, 52(3), 2–22. <https://doi.org/10.1177/002224298805200302>
- Zhang, C., Dong, R., Zhang, X., & Li, W. (2022). A Comparative Analysis of Supervised/unsupervised Algorithms in PHM Application. *2022 Global Reliability and Prognostics and Health Management Conference, PHM-Yantai 2022*. <https://doi.org/10.1109/PHM-Yantai55411.2022.9942049>

8. ANEXOS

Tabela 1: Performance em treino

Modelo	AUC	CA
kNN	0,978	0,918
Logistic Regression	0,968	0,923
Naive Bayes	0,939	0,912
Tree	0,957	0,910

Tabela 2: Performance em teste real

Modelo	AUC	CA
Logistic Regression	0,942	0,902
Tree	0,924	0,903
Naive Bayes	0,935	0,906
kNN	0,922	0,894

Tabela 3: Comparação - treino vs teste

Modelo	AUC Treino	AUC Teste	Queda AUC	CA Treino	CA Teste	Queda CA
Logistic Regression	0,968	0,942	-0,026	0,918	0,902	-0,016
Naive Bayes	0,939	0,935	-0,004	0,923	0,903	-0,02
Tree	0,957	0,924	-0,033	0,912	0,906	-0,006
kNN	0,978	0,922	-0,056	0,910	0,894	-0,016

Tabela 4: ranking geral (AUC + CA)

Modelo	Estabilidade AUC	Estabilidade CA	Ranking Geral
Tree	3° (-0,033)	1° (-0,006)	1°
Logistic Regression	2° (-0,026)	2° (-0,016)	2°
Naive Bayes	1° (-0,004)	4° (-0,020)	3°
kNN	4° (-0,056)	2° (-0,016)	4°

Tabela 5: Segmentação dos rendimentos

Segmentação	Nível rendimento
Básico	Até 800€
Médio-Baixo	801€ a 1.500€
Médio	1.501€ a 2.500€
Médio-Alto	2.501€ a 4.000€
Premium	Acima de 4.000€

Tabela 6: Segmentação Geracional por Faixas Etárias

Idade	Geração
18-22 anos	Z Jovens
23-30 anos	Z Adultos
31-38 anos	Millennials Jovens
39-45 anos	Millennials Maduros
46-53 anos	X Jovens
54-60 anos	X Maduros
61-70 anos	Boomers Ativos

Tabela 7: Variáveis de comportamento relacional

Variáveis comportamentais	Escala			
	Lento	Demorado	Normal	Rápido
TempoProcesso				
ComportamentoESpecialistaComunicacaoEscrita	Pouco Escrita	Escrita Normal	Muito Escrita	
ComportamentoESpecialistaAtenderChamadas	Nada Conversador	Pouco Conversador	Conversador	Muito Conversador
ComportamentoESpecialistaDisponibilidade	Nada Disponível	Pouco Disponível	Disponível	Muito Disponível
ComportamentoPatientComunicacaoEscrita	Pouco Escrita	Escrita Normal	Muito Escrita	
ComportamentoPatientAtenderChamadas	Nada Conversador	Pouco Conversador	Conversador	Muito Conversador

ComportamentoPatientDisponibilidade	NadaDisponível	PoucoDisponível	Disponível	MuitoDisponível
ComportamentoPatientComunicacaoVoz	NadaConversador	PoucoConversador	Conversador	MuitoConversador
ComportamentoPatientTipoInteração	PoucoDigital	Digital	Camaleões	MuitoDigital

Tabela 8: Estrutura de variáveis do *dataset*

Categoria	Elemento	Descrição
Identificação	PatientID	Identificador único do paciente/cliente
	GrupoSourcesA	Grupo de origem ou fonte de dados do cliente
Perfil Socioeconómico	SegmentRendimento	Segmento de rendimento do cliente
	NºDependentes	Número de dependentes do cliente
	SetorAtividade	Setor de atividade profissional do cliente
Perfil Demográfico	SegmentEtárioGeração	Segmento etário/geração
	Genero	Género do cliente (masculino/feminino)
	EstadoCivil	Estado civil do cliente
	EstadoCivil_Filhos	Estado civil considerando a presença de filhos
Segmentação Geográfica	Distrito	Distrito de residência do cliente
Perfil Tecnológico	SegmentCompTecnologico	Segmento de competência tecnológica do cliente
Métricas de Engagement	NumeroInscricoesCH	Número de inscrições em crédito habitação
	NumeroInscricoesCPCC	Número de inscrições em crédito pessoal
	TotaldeInscrições	Total de inscrições em todos os serviços
	Inscrições_Cliente	A ordem como fez as inscrições

Comportamento Relacional	ComportamentoESpecialistaComunicacaoEscrita	Comportamento do especialista na comunicação escrita
	ComportamentoPatientComunicacaoEscrita	Comportamento do cliente na comunicação escrita
	ComportamentoESpecialistaAtenderChamadas	Comportamento do especialista ao atender chamadas
	ComportamentoESpecialistaDisponibilidade	Disponibilidade do especialista
	ComportamentoPatientAtenderChamadas	Comportamento do cliente ao atender chamadas
	ComportamentoPatientDisponibilidade	Disponibilidade do paciente
	ComportamentoPatientComunicacaoVoz	Comportamento do cliente na comunicação por voz
	ComportamentoPatientTipoInteraçao	Tipo de interação preferida pelo cliente
	TempoProcesso	Tempo de duração do processo
Carteira Seguros de	PossuiApoliceVida2024	Possui apólice de seguro de vida em 2024 (sim/não) (variável target)
	PossuiApoliceVidaAntes2024	Possuía apólice de seguro de vida antes de 2024 (sim/não)
	PossuiApoliceAutoAntes2024	Possuía seguro automóvel antes de 2024 (sim/não)
	PossuiApoliceMRAntes2024	Possuía seguro multiriscos antes de 2024 (sim/não)
	PossuiApoliceSaudeAntes2024	Possuía seguro de saúde antes de 2024 (sim/não)
	PossuiApoliceAPAntes2024	Possuía seguro de acidentes pessoais antes de 2024 (sim/não)
	PossuiApoliceRCAntes2024	Possuía seguro de responsabilidade civil antes de 2024 (sim/não)

PossuiApolicePetsAntes2024	Possuía seguro para animais de estimação antes de 2024 (sim/não)
PossuiApoliceATAntes2024	Possuía seguro de assistência em viagem antes de 2024 (sim/não)
PossuiApolicePPRAntes2024	Possuía Plano Poupança Reforma antes de 2024 (sim/não)
PossuiMaisQueUmTipoSeguroAntes2024	Possuía mais de um tipo de seguro antes de 2024 (sim/não)
PossuiApenasUmTipoSeguroAntes2024	Possuía apenas um tipo de seguro antes de 2024 (sim/não)
SemSeguroAntes2024	Não possuía qualquer seguro antes de 2024 (sim/não)

Imagem 1 (origem Orange): Matriz confusão - *Logistic Regression*

		Predicted		Σ
		Não	Sim	
Actual	Não	86.0 %	14.0 %	357
	Sim	5.7 %	94.3 %	367
Σ		328	396	724

Imagem 2 (origem Orange): Matriz confusão - *kNN*

		Predicted		Σ
		Não	Sim	
Actual	Não	84.0 %	16.0 %	357
	Sim	5.4 %	94.6 %	367
Σ		320	404	724

Imagem 3 (origem Orange): Matriz confusão - *Naive Bayes*

		Predicted		Σ
		Não	Sim	
Actual	Não	81.5 %	18.5 %	357
	Sim	0.5 %	99.5 %	367
Σ		293	431	724

Imagem 4 (origem Orange): Matriz confusão - *Tree*

		Predicted		Σ
		Não	Sim	
Actual	Não	84.6 %	15.4 %	357
	Sim	4.1 %	95.9 %	367
Σ		317	407	724

Imagem 5 (origem Orange): Nó terminal da *decison tree* - Árvore para “PossuiApoliceVida2024: Sim”

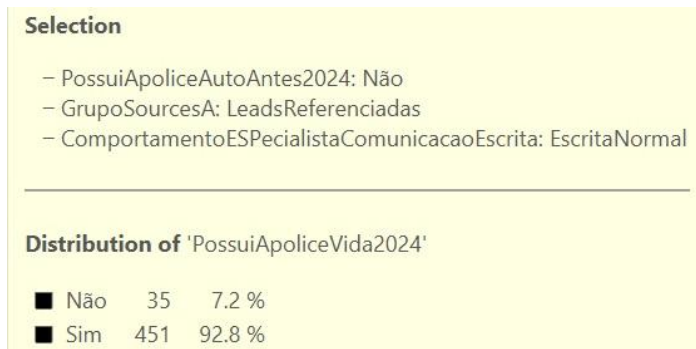


Imagem 6 (origem Orange): Nó terminal da *decison tree* - Árvore para “PossuiApoliceVida2024: Não”

