INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Modelo Preditivo Baseado em Machine Learning para Riscos de Cibersegurança em Projetos Tecnológicos

Fernando João Piedade

Mestrado em Engenharia Informática

Orientador:
Doutorada Luísa Cristina da Graça Pardal Domingues Miranda,
Professora Auxiliar,
Iscte - Instituto Universitário de Lisboa

Orientador:
Doutorada Virgínia Maria da Silva Araújo, Investigadora Associada,
ISTAR-Iscte - Centro de Investigação em Ciências da Informação,
Tecnologias e Arquitetura

Outubro, 2025

TECNOLOGIAS
E ARQUITETURA

Departamento de Ciências e Tecnologias da Informação

Modelo Preditivo Baseado em Machine Learning para Riscos de Cibersegurança em Projetos Tecnológicos

Fernando João Piedade

Mestrado em Engenharia Informática

Orientador:
Doutorada Luísa Cristina da Graça Pardal Domingues Miranda,
Professora Auxiliar,
Iscte - Instituto Universitário de Lisboa

Orientador:
Doutorada Virgínia Maria da Silva Araújo, Investigadora Associada,
ISTAR-Iscte - Centro de Investigação em Ciências da Informação,
Tecnologias e Arquitetura

Outubro, 2025

Dedico esse trabalho a todos pela oportunidade e paciência!

Agradecimento

Expresso meus sinceros agradecimentos a todos que, contribuíram para a concretização deste projeto, e de todo percurso durante o mestrado no Iscte.

Às minhas orientadoras, Prof.ª Dra. Luísa Domingues e a Prof.ª Dra. Virgínia Araújo, deixo um agradecimento especial pela excelência, rigor académico e dedicação com que me acompanharam ao longo de todas as etapas deste trabalho.

Estendo a minha gratidão a todos os professores do curso, pelas valiosas contribuições, ensinamentos e pelo compromisso demonstrado ao longo da formação. Cada disciplina, cada partilha e cada desafio proposto foram fundamentais para a construção do conhecimento aqui aplicado.

Agradeço profundamente à minha família pelo apoio constante, à minha esposa Zelina Piedade, pela paciência, compreensão e incentivo, mesmo nos momentos mais exigentes. Aos meus filhos Kael Fernando Piedade e a Ayla Uriela Piedade, que, mesmo sem entenderem totalmente a natureza deste percurso, sempre respeitaram o meu tempo de dedicação e, com os seus gestos simples, tornaram-se fonte diária de motivação.

Aos meus amigos, aos colegas, pelo companheirismo, troca de experiências e apoio ao longo desta jornada. O convívio académico e pessoal com cada um de vocês foi enriquecedor e marcante. Realçando a força e o encorajamento do meu amigo Nkandi António, que verdadeiramente se tornou aquele amigo-irmão,

A todos, o meu profundo reconhecimento e muito obrigado.

Resumo

A crescente digitalização de processos em diferentes setores acentua a exposição de dados sensíveis, sistemas, recursos e ativos a riscos de cibersegurança, especialmente em projetos tecnológicos. Este trabalho propõe um modelo inteligente para previsão de riscos de cibersegurança, resultante da integração entre abordagens tradicionais de gestão de riscos e técnicas de Machine Learning (ML).

A investigação segue a metodologia *Design Science Research* (DSR), estruturada em seis etapas, desde a identificação do problema até à comunicação dos resultados. No âmbito do rigor científico, foi realizada uma Revisão Sistemática da Literatura (SLR), apoiada pela abordagem PRISMA, permitindo identificar lacunas e tendências na aplicação de modelos preditivos em contextos de cibersegurança.

O trabalho apresenta um protótipo capaz de consumir e processar dados históricos, organizados num *dataset*, para o treino de algoritmos de ML, gerando modelos preditivos aplicados à identificação de riscos de cibersegurança. Adicionalmente, compara três algoritmos de ML, avaliando a sua eficácia e utilidade na previsão de riscos cibernéticos.

A implementação foi validada por meio da avaliação de métricas de desempenho dos algoritmos num ambiente controlado, com dados sintéticos que replicam cenários operacionais realistas, complementada por um *dashboard* interativo que apresenta os resultados de forma clara, demonstrando a viabilidade técnica do protótipo e o seu potencial para antecipar riscos. Recomenda-se, contudo, validação em contextos de produção.

A proposta contribui para o domínio académico ao demonstrar a aplicabilidade de modelos preditivos sobre riscos cibernéticos e evidencia o potencial de integração de ML em processos de gestão de riscos em projetos tecnológicos.

Palavras-chave: *previsão de riscos de cibersegurança, aprendizagem automática para avaliação de riscos, gestão de riscos de projetos, modelação preditiva de segurança, classificação supervisionada de cibersegurança, gestão de dados sintéticos para analise de riscos*.

Abstract

The increasing digitalization of processes in various sectors increases the vulnerability of sensitive data, systems, resources, and assets to cybersecurity risks, especially in technological projects. This work proposes an intelligent model to predict cybersecurity risks, obtained by integrating traditional risk management approaches with Machine Learning (ML) techniques.

The research follows the Design Science Research (DSR) methodology, structured in six stages, from problem identification to results communication. As part of the scientific rigor, a Systematic Literature Review (SLR) was conducted, supported by the PRISMA approach, allowing the identification of gaps and trends in the application of predictive models in cybersecurity contexts.

The work presents a prototype capable of consuming and processing historical data, organized in a dataset, for the training of ML algorithms, generating predictive models applied to the identification of cybersecurity risks. Furthermore, it compares three ML algorithms, evaluating their effectiveness and usefulness in predicting cybersecurity risks.

The implementation was validated by evaluating the algorithms' performance metrics in a controlled environment, using synthetic data that replicates realistic operational scenarios, complemented by an interactive dashboard that clearly presents the results, demonstrating the prototype's technical feasibility and its potential to anticipate risks. However, validation in real-world production contexts is recommended.

This proposal contributes to the academic field by demonstrating the applicability of predictive models to cybersecurity risks and highlighting the potential for integrating machine learning (ML) into risk management processes in technological projects.

Keywords: cybersecurity risk prediction, machine learning for risk assessment, project risk management, predictive security modeling, supervised classification in cybersecurity, synthetic data generation for risk analysis.

Índice

Índice de Tabelas

Índice de Figuras

Índice de Gráficos

Glossário de Abreviaturas e Siglas

IA – Inteligência Artificial
ML – Machine Learning
PM – Project Management
PMI – Project Management Institute
PMBOK – Project Management Body of Knowledge
PRINCE2 – Projects IN Controlled Environments
BMI - Building Information Modeling
PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analysis
NN – Neural Network
LLM – Large Language Model
NLP – Natural Language Processing
SLR – Systematic Literature Review
DSR – Design Science Research
DRL – Deep Reinforcement Learning
CVE — Common Vulnerabilities and Exposures
ICS – Industrial Control System
AUC-ROC - Area Under the Receiver Operating Characteristic Curve

CAPÍTULO 1

Introdução

1.1. Contexto

A crescente complexidade do panorama de ameaças cibernéticas coloca em evidência a importância da cibersegurança nos projetos tecnológicos e na transformação digital [1] [2]. Com a expansão contínua de diferentes formas de ataques, com a introdução de novos ativos, novas interfaces e novas integrações, as vulnerabilidades aumentam [3] [4], levando as organizações a enfrentarem riscos cada vez mais desafiadores para a proteção dos seus dados e sistemas críticos [5] [6].

Nesse cenário, a Inteligência Artificial (IA) e especificamente a área de Machine Learning (ML) emergem como aliados estratégicos [5] [7]. Estudos recentes demonstram que soluções baseadas em IA reduzem o tempo médio de deteção e resposta a incidentes [8], oferecendo não apenas mecanismos reativos, mas também abordagens proativas capazes de identificar padrões e antecipar potenciais riscos [9].

Os frameworks [10], como o NIST Cybersecurity Framework [11], as normas ISO 27005 [12] e ISO 31000 [13], bem como linhas orientadoras com destaque ao PMBOK 6ª e 7ª edição [14][15], forneçam fundamentos para a gestão de riscos [16], tais abordagens ainda dependem fortemente de análise manual e carecem de mecanismos dinâmicos para acompanhar a velocidade das ameaças digitais [9]. Essa lacuna reforça a necessidade de integração de IA/ML no processo de avaliação e previsão de riscos [17].

A cibersegurança, enquanto disciplina, baseia-se na proteção de ativos organizacionais tais como dados, sistemas, processos e pessoas, contra potenciais incidentes [18]. Essa proteção é constantemente desafiada pela existência de vulnerabilidades, que representam fragilidades técnicas ou humanas, e por ameaças, entendidas como agentes ou eventos capazes de explorar tais vulnerabilidades [9]. O risco surge precisamente quando uma ameaça consegue explorar uma vulnerabilidade num ativo, produzindo impacto negativo para a organização e para os projetos em execução [19].

Assim, este trabalho propõe explorar a utilização de técnicas de IA e ML não apenas para a deteção de vulnerabilidades e ameaças, mas principalmente para a análise preditiva de riscos cibernéticos em projetos tecnológicos. O objectivo é oferecer recomendações práticas para a proteção de ativos, fortalecendo a tomada de decisão e estabelecendo uma abordagem eficiente na gestão de riscos de cibersegurança em ambientes digitais.

1.2. Motivação

A gestão de riscos de cibersegurança é essencial nos projetos tecnológicos [16], pois permite antecipar, evitar e mitigar eventos que podem comprometer ativos, dados e sistemas críticos [19]. Contudo, à medida que os projetos se tornam mais complexos [20], os frameworks atuais mostram limitações, uma vez que ainda dependem de análises manuais e oferecem pouca capacidade de previsão dinâmica dos riscos [21].

Nesse contexto, a IA e ML destacam-se ao ampliar a capacidade de deteção e resposta a ameaças [22], com aplicações em análise de vulnerabilidades, comportamento anómalo [23], *threat intelligence* [24] e resposta automatizada [25]. Mais recentemente, essas tecnologias têm possibilitado o desenvolvimento de módulos preditivos, capazes de correlacionar ativos, vulnerabilidades e ameaças [25], antecipando riscos de cibersegurança em tempo real [5].

Perante este cenário, esta dissertação tem no seu cerne o desenvolvimento de um módulo preditivo de riscos de cibersegurança, concebido para apoiar a tomada de decisão e a proteção de ativos no âmbito de projetos tecnológicos. Esse módulo, pode adicionalmente ser complementado por um outro responsável para extrair e estruturar informações de documentos de projeto, recorrendo a *Natural Language Processing* (NLP) [26]. Embora relevante, não constitui o foco central deste trabalho, funcionando apenas como um componente de suporte [27][28].

Com essa abordagem, pretende-se oferecer um mecanismo de monitorização e preditivo de identificação de riscos de cibersegurança em projetos, assim, antecipando potenciais problemas antes que se materializem. A proposta pretende demonstrar como a integração de técnicas de ML pode transformar dados históricos em conhecimento preditivo, orientando decisões de mitigação e contribuindo para uma gestão de riscos mais inteligente e eficiente.

1.3. Objectivos

1.3.1. Objectivo Geral

O objectivo geral da dissertação consiste no desenvolvimento de um módulo de Inteligência Artificial destinado à previsão de riscos de cibersegurança em projetos tecnológicos, através da aplicação de algoritmos de *Machine Learning* sobre dados estruturados, para orientar estratégias de mitigação e apoiar a tomada de decisão ao longo do ciclo de vida dos projetos.

1.3.2. Objectivos Específicos

- **OB1:** Identificar e comparar algoritmos de ML adequados para classificação de riscos cibernéticos em contexto de projetos;
- **OB2:** Conceber uma arquitetura integrando processamento de dados e algoritmos de ML para previsão de riscos de cibersegurança;
- **OB3:** Implementar protótipo como prova-de-conceito da viabilidade técnica.

1.4. Abordagem metodológica

Este trabalho adota a metodologia *Design Science Research* (DSR), amplamente reconhecida na área de Sistemas de Informação pela sua capacidade de guiar o desenvolvimento de artefactos tecnológicos inovadores com base em problemas reais [29].

A DSR organiza-se em seis etapas fundamentais: (1) identificação do problema e motivação; (2) definição dos objetivos da solução; (3) desenho e desenvolvimento do artefacto; (4) demonstração; (5) avaliação; e (6) comunicação. Esta abordagem permite combinar rigor científico com relevância prática, assegurando que os resultados produzidos sejam úteis e aplicáveis em contextos reais [29].

A escolha pela *Design Science Research* (DSR) justifica-se pela natureza aplicada desta investigação, cujo objetivo central é propor e validar um módulo preditivo de riscos de cibersegurança em projetos tecnológicos. DSR orienta o desenvolvimento de artefactos inovadores em sistemas de informação, combinando rigor científico com relevância prática [29]. Complementarmente, aplicou-se uma Revisão Sistemática da Literatura (SLR) nas fases 1 e 2 da DSR e inclusão do método PRISMA, com suporte do diagrama, com o propósito de fundamentar a identificação do problema, caracterizar o estado da arte e definir os objetivos da solução.

Ao longo da dissertação, cada etapa da metodologia é explorada e aplicada de forma sistemática, orientando o processo de construção, avaliação e validação do protótipo proposto.

1.5. Organização da dissertação

Esta dissertação está estruturada em sete capítulos, organizados de forma a permitir uma progressão lógica entre a fundamentação teórica, o desenvolvimento do artefacto proposto e a sua validação prática.

Capítulo 1 – Introdução: Apresenta o contexto da pesquisa, a motivação do estudo, os objetivos gerais e específicos. Este capítulo estabelece as bases que justificam a relevância do trabalho.

Capítulo 2 – Revisão da Literatura: Explora os principais conceitos relacionados com riscos de cibersegurança em projetos, modelos preditivos aplicados à gestão de riscos, ativos digitais e ameaças, e identifica as lacunas existentes na literatura que este trabalho procura colmatar.

Capítulo 3 – Metodologia: Descreve a abordagem metodológica adotada — *Design Science Research* — e detalha cada uma das suas etapas, desde a identificação do problema até à comunicação dos resultados.

Capítulo 4 – Desenvolvimento do modelo preditivo: Detalha o processo de construção do protótipo proposto, incluindo a definição da arquitetura do sistema, a estrutura dos dados de entrada, a geração de dados, o pré-processamento dos dados e o processo de treino e avaliação dos modelos de *Machine Learning* utilizados.

Capítulo 5 – Implementação e validação do protótipo: Apresenta a implementação prática do protótipo desenvolvido, o ambiente de desenvolvimento, a integração dos modelos com uma API e a interface de utilizador. São também descritos os procedimentos experimentais utilizados para validar o sistema, bem como as métricas de avaliação aplicadas.

Capítulo 6 – Análise e discussão dos resultados: Analisa os resultados obtidos a partir dos testes realizados, comparando o desempenho dos algoritmos, discutindo os insights derivados e validando a eficácia da arquitetura preditiva implementada.

Capítulo 7 – Conclusões e trabalhos futuros: Resume as principais contribuições do trabalho, reconhece as limitações da investigação e propõe linhas de continuidade para trabalhos futuros. Este capítulo finaliza a dissertação, destacando o valor acrescentado do protótipo desenvolvido para a área da gestão de riscos em cibersegurança. São também discutidas limitações técnicas e desafios encontrados.

Complementam a dissertação as secções de Referências Bibliográficas, Apêndices, que contêm materiais de apoio, dados adicionais e documentação complementar relevante para a compreensão e reprodutibilidade do estudo.

CAPÍTULO 2

Revisão da literatura

A presente revisão da literatura tem como objetivo analisar o estado atual da investigação sobre gestão de riscos em cibersegurança, com especial enfoque na integração de técnicas de IA. Para além de oferecer uma visão abrangente sobre o tema, esta análise procura identificar os métodos mais eficazes, os desafios recorrentes e as principais lacunas de conhecimento existentes em previsão de riscos de cibersegurança nos projetos tecnológicos. Pretende-se, assim, avaliar e mapear os estudos já desenvolvidos, em particular aqueles que aplicam abordagens de IA à gestão de riscos, estabelecendo uma base conceptual e metodológica sólida que sustente o desenvolvimento da solução proposta neste trabalho.

2.1. Definição de critério da revisão de literatura

Para garantir uma base de referência sólida, adotou-se a metodologia de *Systematic Literature Review* (SLR), reconhecida como adequada para investigações em sistemas de informação e engenharia de software por proporcionar rigor, transparência e reprodutibilidade no levantamento do estado da arte [30]. O processo inicia-se com a definição das questões de investigação, seguido pela especificação das palavras-chave e dos critérios de inclusão e exclusão, tal como recomendado em estudos recentes que aplicam SLR em projetos tecnológicos e de cibersegurança [1] [31].

A seleção das fontes de dados prioriza bases de dados académicas, como a SCOPUS e a *Web of Science* (WoS). A partir disso, realizamos consultas estruturadas (*queries*) adaptadas a cada base de dados, para a recuperação de referências relevantes e alinhadas ao escopo da dissertação. A revisão também inclui um processo de avaliação e ajuste contínuo para assegurar que os estudos selecionados reflitam as tendências e desafios mais atuais na interseção entre gestão de riscos e Inteligência Artificial.

2.1.1. Perguntas de Pesquisa

As perguntas-chave que sustentaram a pesquisa foram reformuladas para focar especificamente em cibersegurança:

- **RQ1** - Como técnicas de IA e ML podem predizer riscos de cibersegurança durante o desenvolvimento de projetos tecnológicos?
- **RQ2** - Que abordagens existem para integração de ML e avaliação automatizada de riscos de cibersegurança?

- **RQ3** - Como avaliar a eficácia de sistemas de IA na predição de riscos de cibersegurança em projetos reais e quais métricas são mais apropriadas?
- **RQ4** - Como correlacionar efetivamente ativos, vulnerabilidades, ameaças e impactos de negócio para predição de riscos emergentes de cibersegurança?

2.1.2. Base de dados de referências e *keywords*

A seleção das fontes priorizou bases de dados acadêmicas como SCOPUS e *Web of Science* (WoS), com consultas estruturadas adaptadas especificamente para cibersegurança conforme apresentado nas Tabela 1 e Tabela 2.

Keywords principais: "*cybersecurity prediction*", "*AI security assessment*", "*vulnerability prediction*", "*threat modeling automation*", "*project security risks*"

Keywords secundárias: "*automated risk assessment*", *"ML security analysis*", "*cyber threat intelligence*"

Contexto: "*security by design*", "*secure development lifecycle*", "*project cybersecurity*"

As limitações definidas mantiveram-se no período 2020-2025, artigos e revisões em inglês e português, focando especificamente em aplicações de IA para cibersegurança.

Tabela 1 - Números de artigos encontrados na pesquisa na Scopus

SCOPUS			
ALL (("cybersecurity prediction" OR "AI security assessment" OR "vulnerability prediction" OR "threat modeling automation" OR "project security risks") AND ("automated risk assessment" OR "ML security analysis" OR "cyber threat intelligence") OR ("security by design" OR "secure development lifecycle" OR "project cybersecurity")) AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "re"))			
Limitations	**Concept**	**Population**	**Context**
2020-2025	"cybersecurity prediction"	"automated risk assessment"	"security by design"
Only articles and reviews	"AI security assessment"	"ML security analysis"	"secure development lifecycle"
Only Language Eng and Pt	"vulnerability prediction"	"cyber threat intelligence"	"project cybersecurity"
	"threat modeling automation"		
	"project security risks"		
1,963 documents			
50 documents			
46 documents			

Tabela 2 - Números de artigos encontrados na pesquisa na WoS

Web of Science			
ALL (("cybersecurity prediction" OR "AI security assessment" OR "vulnerability prediction" OR "threat modeling automation" OR "project security risks") AND ("automated risk assessment" OR "ML security analysis" OR "cyber threat intelligence") OR ("security by design" OR "secure development lifecycle" OR "project cybersecurity"))			
Limitations	**Concept**	**Population**	**Context**
2020-2025	"cybersecurity prediction"	"automated risk assessment"	"security by design"
Only articles and reviews	"AI security assessment"	"ML security analysis"	"secure development lifecycle"
Only Language Eng and Pt	"vulnerability prediction"	"cyber threat intelligence"	"project cybersecurity"
	"threat modeling automation"		
	"project security risks"		
115 documents			
99 documents			
29 documents			

Para uma melhor visão o gráfico abaixo apresenta o *trend* de artigos encontrados relacionados as principais *keywords* nos últimos 10 anos, com isso aumenta o grau de importância e convicção em haver estudos e soluções que cubram essa temática.

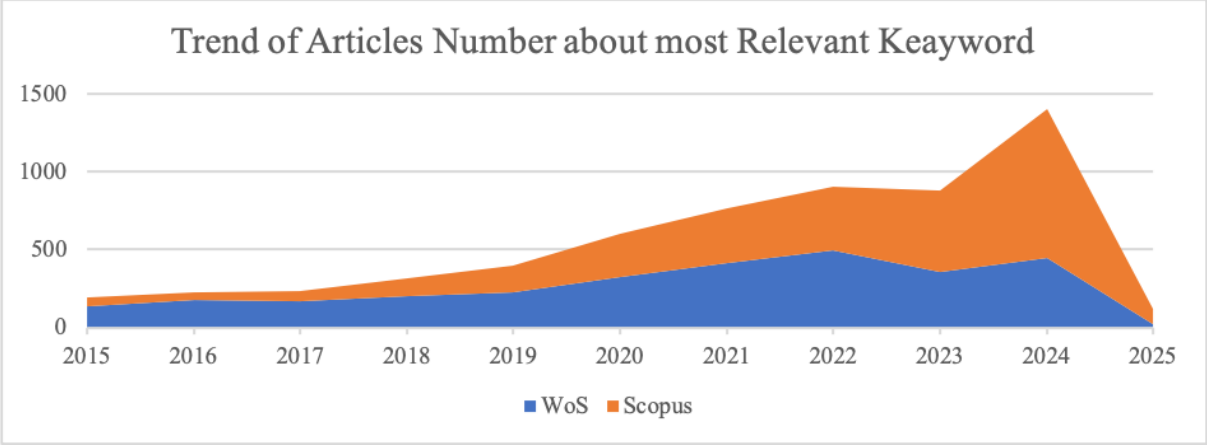


Gráfico 1 - Evolução do número de artigos sobre as Keywords principais

2.1.3. Aplicação do PRISMA no estado da arte

A análise foi conduzida com base nas diretrizes do PRISMA, que garantem transparência no processo de seleção dos estudos. O diagrama de fluxo apresentado na Figura 1 ilustra as etapas do processo: identificação dos estudos nas bases de dados, remoção de duplicados e registos irrelevantes, triagem de acordo com critérios pré-definidos, e inclusão final dos artigos elegíveis. Em cada fase do processo são reportados os números de artigos descartados, evidenciando, por exemplo, duplicações ou ausência de relação direta com o tema.

Na Figura 1, cada passo identificação, triagem, inclusão e exclusão, apresenta quantos artigos foram descartados por exemplo duplicação, ausência de relação com tema, e esse processo, foi na base de análise de tema de cada artigo, análise de cada *abstract* e no final a análise de cada documento.

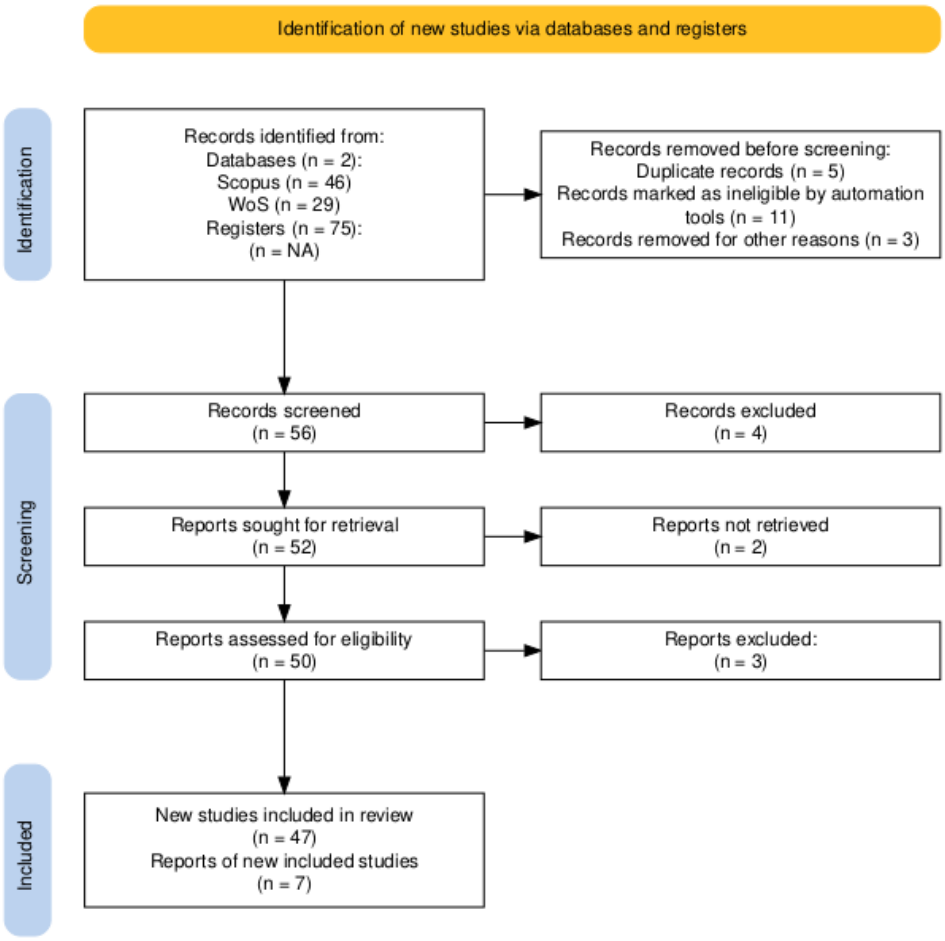


Figura 1 - Fluxograma PRISMA

2.1.4. Análise de literatura com VOSviewer

A pesquisa foi validada usando VOSviewer que mostra as *keywords*, como se relacionam, e as mais relevantes, conforme apresentado na Figura 2, verificando-se que "*cybersecurity*", "*machine learning*" e "*software security*" tiveram maior aparição nos artigos, confirmando uma pesquisa bem orientada.

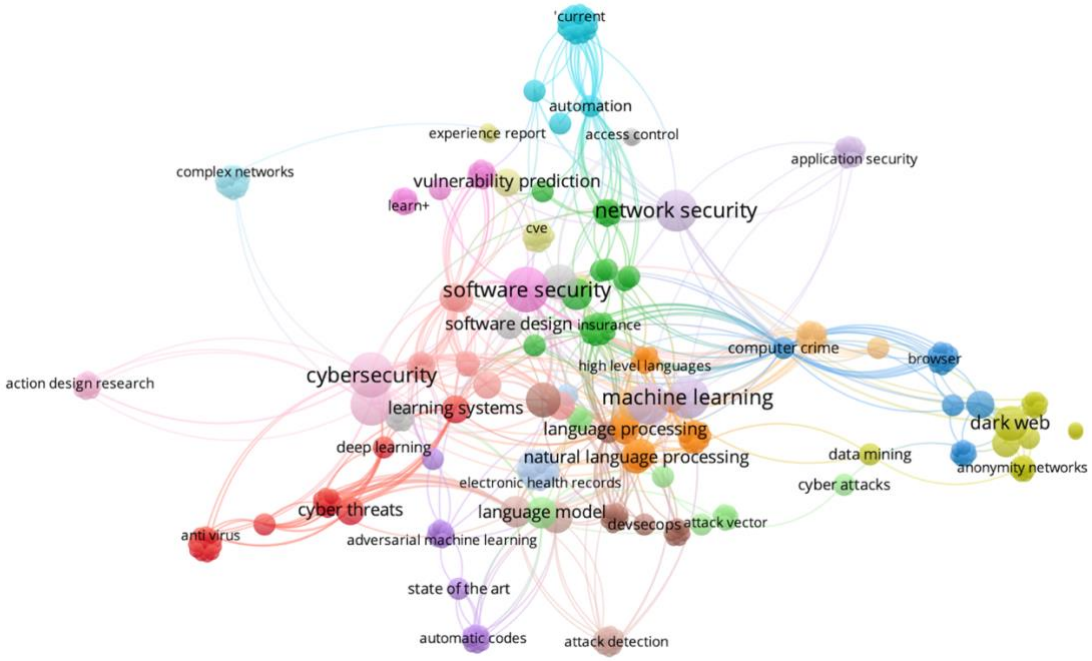


Figura 2 - Tópicos relacionados na pesquisa

2.2. Gestão de Projeto e Riscos

Na literatura, um projeto é geralmente entendido como um empreendimento temporário com objetivos bem definidos, cuja execução exige planeamento, monitorização e controlo sistemático [14]. O PMBOK (7ª edição) define risco como "*um evento ou condição incerta que, se ocorrer, provocará um efeito positivo ou negativo em um ou mais objetivos do projeto*" [15]. Dessa forma, a gestão de riscos surge como um processo essencial para o sucesso de qualquer projeto, envolvendo a identificação, análise, avaliação e mitigação de incertezas que possam comprometer os seus resultados, conforme a Figura 3 adaptado do PMBOK.

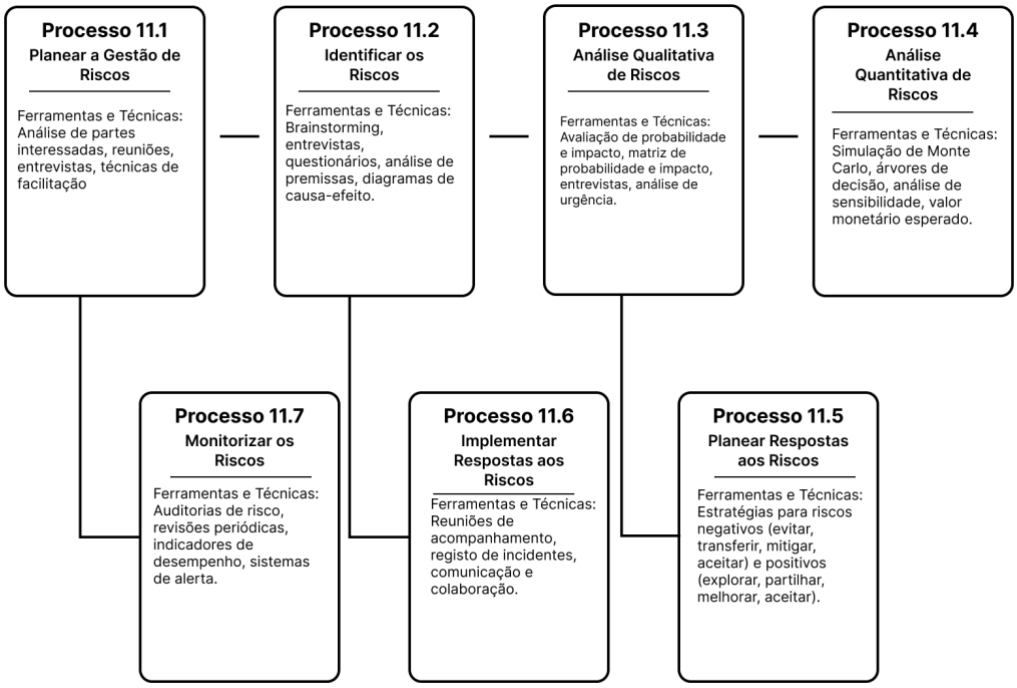


Figura 3 - Processos na gestão de riscos adaptado do PMBOK (6ª edição) [14]

Diversos *frameworks* e normas internacionais oferecem suporte a gestão de riscos, resumida na Tabela 3. O PMBOK (6ª edição), por exemplo, estrutura a gestão de riscos em sete processos que vão do planeamento à monitorização contínua [14]. A norma ISO 31000 fornece princípios transversais aplicáveis a qualquer setor, enfatizando a integração da gestão de riscos em todas as decisões organizacionais [13]. Já a ISO 27005 aborda especificamente riscos relacionados com a segurança da informação, orientando a identificação de ameaças e vulnerabilidades que possam comprometer a confidencialidade, integridade e disponibilidade dos dados [12].

Tabela 3 - Relevância da Normas e Framework para gestão de Risco de Cibersegurança

Norma/Framework	Foco Principal	Relevância para Cibersegurança
PMBOK (6ª/7ª ed.) [14][15]	Gestão de projetos em todas as suas áreas de conhecimento, incluindo riscos, estruturada em processos e boas práticas.	Define o risco como uma incerteza que pode afetar os objetivos do projeto. Serve de base para integrar os riscos cibernéticos no ciclo de vida dos projetos.
ISO 31000, [13]	Princípios e diretrizes gerais para a gestão de riscos, aplicáveis a qualquer setor ou contexto organizacional.	Fornece uma estrutura transversal que pode ser adaptada à cibersegurança, reforçando a importância de uma abordagem sistemática, contínua e integrada da gestão de riscos.
ISO/IEC 27005, [12]	Gestão de riscos de segurança da informação, com ênfase na confidencialidade, integridade e disponibilidade da informação.	Fornece orientações específicas para identificar, analisar, avaliar e tratar riscos de segurança da informação, diretamente relacionados com vulnerabilidades, ameaças e impactos sobre os ativos digitais.

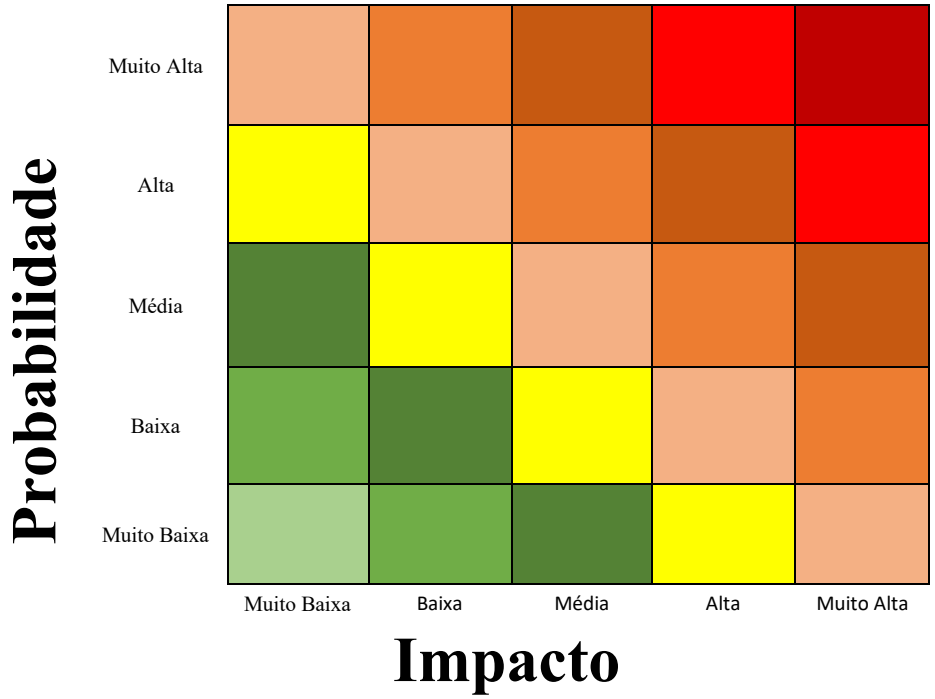
Para melhor visualização há um tratamento dos *outputs*, os estudos analisados organizam as informações de risco por meio de *outputs* visuais (lista de registos, *dashboards*, matriz de riscos, lista de riscos) que facilitam a interpretação e a resposta rápida [18], pois, assim orientam as normas e *frameworks* de gestão de riscos.

Apelando a colorização a matriz de riscos associa a probabilidade, impacto do risco e segmentá-los de forma qualitativa ou quantitativa conforme as Tabela 4 e Tabela 5:

Tabela 4 - Probabilidade e Impacto Risco adaptado PMBOK (6ª edição) [14]

Probabilidade		**Impacto**	
Qualitativa	**Quantitativa**		**Descrição**
Muito Baixa	0,05 (5%)	< 1%	Pequena alteração, quase imperceptível
Baixa	0,10 (10%)	1% – 5%	Ajuste menor, sem afetar entregas-chave
Média	0,30 (30%)	5% – 10%	Ajuste moderado, pode exigir replaneamento
Alta	0,60 (60%)	10% – 20%	Impacto significativo no escopo ou entregas
Muito Alta	0,90 (90%)	> 20%	Compromete fortemente o sucesso do projeto

Tabela 5 - Distribuição com cores da Matriz de Risco adaptado PMBOK (6ª edição) [14]



No campo da cibersegurança, a gestão de riscos ganha contornos específicos [11]. Aqui, os ativos organizacionais — que incluem componentes tangíveis (infraestruturas, sistemas, redes e dispositivos) e não tangíveis (dados sensíveis, propriedade intelectual, reputação e confiança) — tornam-se o centro da análise [12] [13]. Os riscos emergem da possibilidade de uma ameaça explorar uma vulnerabilidade existente nesses ativos, produzindo impacto negativo nos objectivos do projeto e na resiliência organizacional [18].

Dessa forma, a gestão de riscos fornece a metodologia estruturada, enquanto a cibersegurança aplica essa metodologia a um domínio em constante evolução, onde ameaças se transformam rapidamente e exigem abordagens mais adaptativas, dinâmicas e, cada vez mais, preditivas [19].

2.3. Ativos digitais e ameaças em Cibersegurança

No contexto atual de interconexão digital, os ativos digitais constituem a base do funcionamento organizacional e são alvo privilegiado de ameaças cibernéticas. A literatura distingue entre ativos tangíveis, como infraestruturas tecnológicas, sistemas de informação e dispositivos, e ativos não tangíveis, como dados, propriedade intelectual, reputação e confiança organizacional (ISO 31000, 2018 [13]; ISO/IEC 27005, 2022 [12]). Esta distinção é crítica, pois cada tipo de ativo apresenta diferentes vulnerabilidades e implica estratégias diferenciadas de proteção e gestão de riscos.

2.3.1. Ativos tangíveis

Sistemas de Informação: englobam hardware, software e redes responsáveis pelo armazenamento, processamento e transmissão de dados [11]. A sua criticidade advém da dependência operacional diária e da elevada exposição a ataques de *malware*, acessos não autorizados ou exploração de vulnerabilidades [12].

Infraestrutura de Rede: servidores, *routers* e *switches* asseguram a circulação da informação [9]. A interrupção destes ativos, por exemplo através de ataques de negação de serviço (DDoS), pode comprometer a continuidade do negócio [11].

Dispositivos e e*ndpoints*: estações de trabalho, portáteis, tablets e dispositivos móveis são frequentemente portas de entrada para atacantes [9]. A sua proteção é essencial para evitar movimentos laterais em redes corporativas [12].

2.3.2. Ativos não tangíveis

Propriedade Intelectual (IP): patentes, marcas, algoritmos e segredos industriais constituem ativos estratégicos [12]. A perda destes, seja por espionagem industrial ou por insiders maliciosos, representa não apenas perda financeira, mas também vantagem competitiva [13].

Dados: se pode incluir dados pessoais, sensíveis ou estratégicos [12]. A perda, corrupção ou divulgação não autorizada destes pode gerar sanções legais, perdas financeiras e danos reputacionais [15].

Reputação e Marca: o capital reputacional é um ativo não tangível altamente vulnerável a incidentes de cibersegurança [13]. Violações de dados ou falhas públicas em sistemas críticos podem abalar a confiança dos clientes e investidores, comprometendo a sustentabilidade da organização [15], [23].

2.3.3. Ameaças e Riscos Associados a ativos

Cada ativo está exposto a um conjunto de ameaças específicas, que podem resultar em riscos de diferentes níveis de criticidade. A Tabela 6 sintetiza esta relação adaptado de NIST Cybersecurity Framework, ISO e MITRE ATTACK Framework:

Tabela 6 - Relação ativos e ameaças

Ativos	Valor	Principais Ameaças	Riscos Potenciais
Sistemas de Informação [11] , [12]	Alto	*Malware*, exploração de vulnerabilidades, acessos não autorizados [39], [43]	Perda de dados, interrupção de serviços, violação de confidencialidade [13], [16]
Infraestrutura de Rede [12], [39]	Médio	Intrusões, DDoS, falhas de configuração [37], [43]	Interrupção de serviços, perdas financeiras, caos operacional [5], [13] ,[16]
Dispositivos e Endpoints [11], [39]	Médio	*Phishing*, *malware*, roubo de credenciais [19], [43]	Comprometimento de contas, movimentação lateral, violação de dados [13], [16], [24]
Propriedade Intelectual [5], [13]	Alto	Espionagem industrial, insiders [24], [33]	Perda de vantagem competitiva, litígios, estagnação de inovação [5], [16]
Dados [12], [16]	Alto	Infiltração, corrupção, *ransomware* [19], [43]	Multas regulatórias, danos financeiros, perda de confiança [5], [13], [16]
Reputação e Marca [13], [16]	Alto	Divulgação de incidentes, má gestão de crises [24], [33]	Desvalorização da marca, perda de clientes, quebra de confiança [13], [16]

2.4. Critérios de Impacto e Probabilidade em cibersegurança

A avaliação dos riscos associados aos ativos digitais deve basear-se nos princípios fundamentais da Confidencialidade (C), Integridade (I), Disponibilidade (D) e Conformidade (Cf), amplamente reconhecidos como os pilares da segurança da informação [12]. Estes critérios, frequentemente designados pelo modelo CIA(+Cf), constituem a estrutura central para a definição de impacto e probabilidade nas análises de risco em cibersegurança [12].

Confidencialidade (*Confidentiality*) refere-se à protecção da informação contra acessos ou divulgações não autorizadas, assegurando que apenas entidades legítimas possam aceder aos dados. A quebra deste princípio pode comprometer segredos industriais, dados pessoais ou propriedade intelectual, originando consequências legais e reputacionais [12].

Integridade (*Integrity*) diz respeito à exactidão, consistência e completude da informação, prevenindo alterações não autorizadas, quer por erro humano, quer por acções maliciosas. A violação deste princípio pode conduzir à manipulação de relatórios, corrupção de bases de dados ou falhas em processos críticos [12].

Disponibilidade (*Availability*) assegura que os sistemas e os dados estejam acessíveis e funcionais sempre que necessários. Ataques como *Denial of Service* (DoS/DDoS), falhas de infra-estrutura ou interrupções não planeadas representam riscos diretos para este domínio.

Conformidade (*Compliance*), por sua vez, expande o modelo tradicional CIA, refletindo a necessidade de aderir a normas e regulamentos como a ISO/IEC 27005, o Regulamento Geral sobre a Proteção de Dados (RGPD) e o NIST *Cybersecurity Framework* [11]. O incumprimento destes referenciais pode resultar em sanções legais, perdas financeiras e quebra de confiança institucional.

Deste modo, a combinação entre o impacto e a probabilidade de ocorrência, avaliada sob a ótica do modelo CIA(+Cf), fornece uma base quantitativa e qualitativa para a priorização dos riscos e para a definição de estratégias de mitigação alinhadas com as normas internacionais de segurança da informação [11], [12], [13].

2.5. Riscos de Cibersegurança em Projetos

A literatura distingue entre riscos operacionais de cibersegurança, relacionados ao funcionamento diário das infraestruturas digitais, e riscos específicos do ciclo de vida de projetos tecnológicos, associados à introdução de novos ativos, requisitos, integrações e dependências [32]. Estes últimos são particularmente críticos, pois ampliam a superfície de ataque e expõem vulnerabilidades em diferentes fases do desenvolvimento, implementação e manutenção [32][33].

As ameaças cibernéticas exploram vulnerabilidades técnicas, organizacionais e humanas, colocando em risco tanto ativos tangíveis quanto não tangíveis [40]. [25] evidenciam como vulnerabilidades em aplicações móveis podem ser rapidamente exploradas, enquanto [36] destacam a dependência da eficácia dos modelos de deteção da qualidade e diversidade dos *datasets* disponíveis. Em paralelo, estudos [37] e [38] demonstram que ataques em sistemas ciberfísicos e redes inteligentes frequentemente ocorrem em múltiplas etapas [41] [42], reforçando a necessidade de modelação proativa dos riscos.

Durante o ciclo de vida de um projeto, a superfície de ataque evolui de forma dinâmica devido à adição de componentes, interfaces e integrações [18][19]. Esta expansão segue padrões associados a decisões arquitetónicas, práticas de desenvolvimento e requisitos regulatórios. [39] demonstram como plataformas de *Internet of Things* (IoT) ampliam significativamente as vulnerabilidades, enquanto [48] destaca a complexidade acrescida das redes heterogéneas, que exigem novas formas de monitorização para antecipar riscos emergentes.

No que respeita aos *frameworks* de cibersegurança, diretrizes como o NIST Cybersecurity Framework [11], a ISO 27005 [12] e a matriz MITRE ATT&CK [43] oferecem referenciais sólidos para identificar ameaças e definir medidas de mitigação. No entanto, estudos recentes salientam que, embora valiosos, estes referenciais requerem adaptação ao contexto dinâmico dos projetos, sob pena de se tornarem insuficientes face à velocidade e sofisticação das ameaças digitais [44] [33] nesse sentido, a literatura aponta a integração de técnicas de IA e ML como um complemento fundamental, permitindo não só automatizar tarefas rotineiras — como a análise de *logs* e a classificação de alertas [45] [46] — mas, também introduzir uma dimensão preditiva, capaz de antecipar riscos antes da sua materialização.

Assim, os riscos de cibersegurança em projetos tecnológicos não se resumem a incidentes pontuais, mas constituem um processo contínuo e dinâmico que exige gestão proativa e preditiva, alinhada tanto às boas práticas de *frameworks* normativos quanto às capacidades proporcionadas pela IA [19].

2.6. Modelos preditivos na gestão de risco de Cibersegurança e trabalhos relacionados

Com os conceitos acima interiorizados, aqui nessa sessão é apresentado como os modelos preditivos podem ser usados na gestão de riscos, numa abordagem com exemplos de projetos práticos que implementaram algoritmos de ML.

Modelos preditivos tem como base algoritmos de ML, no universo da IA, constatou-se na literatura uma diversidade de algoritmos usados para análise preditiva de riscos com IA, como os apresentados na Tabela 7.

Tabela 7 - Algoritmos de ML/IA aplicados à previsão de riscos de cibersegurança

Categoria	Algoritmo	Estudo(s)	Aplicação / Contribuição
Clássicos	Decision Tree (DT)	Abdullah et al. (2022) [25]	Avaliação de riscos de vulnerabilidades em apps Android; priorização por probabilidade/impacto.
	Random Forest (RF)	Abdullah et al. (2022) [25]; Siewruk & Mazurczyk (2021) [45],	Classificação de vulnerabilidades e deteção de anomalias; robustez em datasets ruidosos.
	Support Vector Machine (SVM)	Siewruk & Mazurczyk (2021) [45]	Classificação supervisionada de vulnerabilidades em software.
Deep Learning (DL)	Redes Neuronais Profundas (DNN)	Arikan et al. (2024) [46]	Deteção de vulnerabilidades em software fechado; automação de threat intelligence.
	Híbrido DL + análise estática	Nithya et al. (2024) [47]	Deteção de ataques de validação de entrada; reforço de segurança preventiva.
	Federated Learning (não supervisionado)	Bertoli et al. (2023) [48]	Deteção de intrusões em redes heterogéneas distribuídas.
Probabilísticos	Modelos context-aware probabilísticos	Noor et al. (2023) [37]	Deteção de ameaças em sistemas ciberfísicos via correlação probabilística.
	Correlação contextual multi-etapas	Sen et al. (2022) [38]	Previsão de cadeias de ataque em redes elétricas inteligentes.
Ensemble / Boosting	XGBoost, LightGBM	Guo et al. (2024) [37]	Eficazes para datasets desbalanceados em vulnerabilidade e intrusão.
Emergentes	Graph Neural Networks (GNN)	Patel et al. (2024) [56]	Modelação de relações entre ativos, vulnerabilidades e ameaças.
	Large Language Models (LLMs)	Scholz et al. (2024) [53]	Potencial em threat intelligence explicativa; desafios de viés e confiabilidade.

A utilização de algoritmos de ML e IA para previsão e mitigação de riscos de cibersegurança tem sido amplamente discutida na literatura recente, com aplicações que vão desde a análise de vulnerabilidades até à deteção de ataques sofisticados em tempo real.

Algoritmos clássicos como *Decision Trees* (DT) e *Random Forest* (RF) continuam a ser amplamente aplicados devido à sua interoperabilidade e robustez. [25] demonstraram a aplicação destes modelos na avaliação de riscos de vulnerabilidades em aplicações Android, destacando a capacidade de priorizar riscos com base em métricas de probabilidade e impacto. [45] também exploraram classificadores supervisionados como RF e *Support Vector Machines* (SVM) para classificação de vulnerabilidades de software, evidenciando melhorias face a abordagens manuais.

Com o crescimento da complexidade das ameaças e dos ambientes digitais, surgiram abordagens baseadas em *Deep Learning* (DL). [46]) propuseram um sistema de deteção de vulnerabilidades em software fechado utilizando técnicas de redes neuronais profundas, capaz de automatizar a produção de inteligência de ameaças partilhável. Da mesma forma, [47] aplicaram técnicas híbridas de análise estática e ML para deteção de ataques de validação de entrada, reforçando a aplicabilidade de modelos preditivos na antecipação de vulnerabilidades. Em contextos mais complexos, como infraestruturas críticas e redes heterogéneas, [48] investigaram a aplicação de *federated learning* com modelos não supervisionados para deteção de intrusões, mostrando o potencial do DL em cenários distribuídos.

Modelos probabilísticos também recebem destaque. Trabalhos como os de [37] propõem arquiteturas *context-aware* para deteção de ameaças em sistemas ciberfísicos inteligentes, utilizando correlação probabilística entre eventos para antecipar riscos. [38] aplicaram correlação contextual multi-etapas em redes elétricas inteligentes, evidenciando que modelos probabilísticos permitem prever cadeias de ataque antes da materialização completa do incidente.

Nos últimos anos, há uma clara ascensão de técnicas ensemble e *boosting*, com destaque para XGBoost e LightGBM, amplamente referidas em estudos recentes de vulnerabilidade e deteção de intrusões. [36], numa análise sistemática de *datasets* de vulnerabilidades [48], destacaram que os modelos *ensemble* são particularmente eficazes para lidar com *datasets* desbalanceados, característicos do domínio da cibersegurança, onde registos de ataques são menos frequentes em comparação ao tráfego legítimo [45].

Para além destas, tendências emergentes incluem o uso de *Graph Neural Networks* (GNNs), capazes de modelar relações entre ativos [19], vulnerabilidades e ameaças em grafos, proporcionando uma visão relacional mais rica sobre riscos cibernéticos [28] [46]. Embora a literatura ainda seja incipiente neste campo, estudos como o de [56] sobre *threat intelligence feeds* apontam para a integração futura de GNNs em pipelines de deteção e previsão. Paralelamente, [59] discutem o impacto de *Large Language Models* (LLMs) na segurança, abrindo caminho para modelos preditivos com capacidades explicativas, embora ainda enfrentem desafios de confiabilidade e viés.

Para uma análise aprofundada compilou-se a Tabela 7, mostra o relacionamento entre algoritmos ML e cibersegurança.

2.6.1. Dados e Feature Engineering

Um aspeto crítico na aplicação de algoritmos de ML em cibersegurança é o *feature engineering*, isto é, o processo de seleção, criação e transformação das variáveis que melhor traduzem as características relevantes dos riscos, das vulnerabilidades e dos ativos sob análise. [34] [35]. A literatura destaca que a sustentabilidade dos modelos não depende apenas do algoritmo escolhido, mas da qualidade e diversidade das características dos dados utilizados [36].

A Tabela 8 apresenta exemplos de dados de entrada (*features*) identificados na revisão de literatura. [25], por exemplo, utilizaram métricas de permissões, configurações e código-fonte em aplicações Android para classificar vulnerabilidades. Já [47] combinaram *logs* de execução e parâmetros de entrada para detetar ataques de validação, enquanto [48] mostraram como dados de tráfego de rede distribuído podem alimentar modelos federados para deteção de intrusões em redes heterogéneas. Em paralelo, [37] exploraram variáveis contextuais em sistemas ciberfísicos — como sensores de desempenho, eventos temporais e estados do sistema — aplicando modelos probabilísticos para antecipar cadeias de ataque.

Tabela 8 - Tipos de dados de entrada

Referência	Contexto	Algoritmo(s)	Dados de entrada (features)
Abdullah et al. (2022) [25]	Riscos em aplicações Android	Decision Tree, Random Forest	Permissões, APIs chamadas, código-fonte; treino/teste 70/30
Nithya et al. (2024) [47]	Deteção de ataques de input validation	Híbrido DL + análise estática	Logs de execução, parâmetros de entrada, fluxos de chamadas
Bertoli et al. (2023) [48]	Redes heterogéneas distribuídas	Federated Learning (não supervisionado)	Tráfego de rede, métricas de latência, pacotes suspeitos
Noor et al. (2023) [37]	Sistemas ciberfísicos	Modelos probabilísticos context-aware	Sensores, estados do sistema, eventos temporais, correlação entre logs
Arikan et al. (2024) [46]	Vulnerabilidades em software fechado	Redes neuronais profundas	Reportes de falhas, descrições textuais (NLP), métricas de execução

Divisão dos Dados e Validação: A literatura indica que a prática comum é dividir os dados em treino (60–80%), teste (20–40%) e, em alguns casos, validação cruzada (k-fold) para aumentar a robustez [48][25]. Em contextos temporais, como previsão de ataques em redes inteligentes, [37] utilizaram validação temporal para garantir consistência.

2.6.2. Métricas de Avaliação segundo a Literatura

Nos trabalhos analisados sobre riscos de cibersegurança, a avaliação dos modelos concentra-se sobretudo em métricas de classificação, já que o objetivo central é distinguir entre tráfego legítimo e malicioso, ou ainda identificar vulnerabilidades exploráveis. Entre as métricas mais utilizadas destacam-se:

- Exatidão (*Acurrecy*): mede a proporção de previsões corretas face ao total de instâncias analisadas [45];
- Precisão (*Precision*): indica a percentagem de previsões positivas que são efetivamente corretas, essencial para reduzir falsos positivos [17];
- Sensibilidade (*Recall*): avalia a capacidade do modelo em identificar todos os casos relevantes (verdadeiros positivos) [17];
- F1-score: combina precisão e *recall* numa média harmónica, útil em cenários de classes desbalanceadas, típicos em *datasets* de cibersegurança [48];

- AUC-ROC: métrica frequente para avaliar a capacidade discriminatória dos modelos, especialmente em deteção de intrusões [48].

Estudos como os de [25][45] evidenciam o uso de *precision*, *recall* e F1-score como indicadores-chave na classificação de vulnerabilidades. Já [36] salientam que *datasets* de vulnerabilidade frequentemente apresentam classes desbalanceadas, o que torna o F1-score e a AUC-ROC métricas mais robustas. Trabalhos recentes, como os de [48][46], reforçam esta tendência ao aplicar métricas de classificação em cenários de redes heterogéneas e deteção de vulnerabilidades em software fechado, respetivamente, evidenciando a necessidade de reduzir tanto falsos positivos como falsos negativos.

A Tabela 9 aponta para todas as referências com o tema, área de aplicação e comentários centrados nos algoritmos de ML, compartilham o foco na integração de métodos de IA com estratégias de gestão de riscos de cibersegurança, que é o cerne da dissertação, digamos que aqui encontra-se concentrada todas as informações relacionadas com o projeto.

Tabela 9 - Referências e áreas aplicadas e algoritmos de ML

Nº	Referência	Área / Aplicação	Algoritmo(s) / Abordagem	Dados / Features	Métricas usadas	Comentários e Limitações
1	Abdullah et al., 2022, [25]	Mobile / Apps Android	Árvores de decisão, Random Forest (classificação de vulnerabilidades / risco)	Permissões, APIs chamadas, artefactos de código/manifest	Acurácia, Precisão, Recall, F1	Quantifica e prioriza riscos em apps móveis, mas depende fortemente da qualidade dos metadados e carece de validação em ambientes reais de distribuição de aplicações.
2	Siewruk & Mazurczyk, 2021 [45]	Segurança de software	RF, SVM (classificação context-aware de vulnerabilidades)	Metadados de CVEs, contexto do software	Acurácia, F1, AUC	Reduz falsos positivos via contexto, porém apresenta limitações na generalização para diferentes domínios e versões de software.
3	Guo et al., 2024, [36]	Datasets de vulnerabilidade	Estudo/survey (ênfase em desbalanceamento e qualidade)	Catálogos de datasets públicos	—	Evidencia desbalanceamento e baixa padronização dos datasets; recomenda F1/AUC sobre acurácia. Limitação: ausência de propostas experimentais diretas.
4	Arikan et al., 2024, [46]	CTI / software fechado	Deep Learning para deteção de vulnerabilidades e CTI partilhável	Descrições de vulnerabilidades, artefactos binários/texto	F1, Recall	Automatiza CTI, mas sofre de falta de interpretabilidade e alto custo computacional. Dificuldade de auditoria e replicabilidade.
5	Nithya et al., 2024, [47]	Web / Input validation	Híbrido análise estática + ML	Logs de execução, parâmetros de entrada	Acurácia, F1	Melhora deteção precoce, mas requer amostras amplas e sofre de overfitting em ambientes web heterogéneos.
6	Bertoli et al., 2023, [48]	IDS em redes heterogéneas	Federated learning não supervisionado/empilhado	Tráfego/fluxos de rede distribuídos	Acurácia, AUC	Reduz necessidade de partilha de dados, mas enfrenta problemas de latência e coordenação entre nós federados.
7	Noor et al., 2023, [37]	CPS inteligentes	Context-aware (probabilístico) para deteção e resposta	Eventos/sensores, correlações temporais	Precisão, Recall, F1	Antecipação de ataques multi-etapas, mas desempenho

						dependente da sincronização e granularidade temporal dos dados.
8	Sen et al., 2022, [38]	Smart grids	Correlação contextual multi-estágio	Telemetria de rede/energia, eventos	Precisão, Recall, F1	Previsão de cadeias de ataque eficaz, porém difícil de escalar para redes maiores e ambientes distribuídos.
9	Fortino et al., 2022, [39]	IoT / plataformas	Análise comparativa de plataformas (segurança)	Catálogo de controlos e integrações	—	Evidencia lacunas de integração, mas sem avaliação quantitativa. Limitação: dependência de análises descritivas.
10	Arief et al., 2020, [32]	ICS / processo industrial	Segmentação de ICS (arquitetural)	Topologias ICS, zonas/conduítes	—	Reduz efeito dominó em ICS, porém carece de avaliação prática em sistemas industriais reais.
11	Feidakis et al., 2021, [49]	Infraestruturas críticas	Plataforma integrada (projeto DESMOS)	Sensores/IoT, dados operacionais	—	Integra segurança e monitorização, mas ainda experimental; escalabilidade limitada.
12	Rueda-Rueda & Portocarrero, 2021, [50]	IoT / frameworks	Revisão de medidas de segurança	Catálogo de controlos e boas práticas	—	Sistematiza boas práticas, mas não valida empiricamente o impacto das medidas propostas.
13	Granata & Rak, 2024, [44]	Threat modeling	Comparação de ferramentas OSS (automação)	Modelos de arquitetura/ameaças	—	Automatiza modelagem de ameaças, mas cobertura limitada a certos frameworks (STRIDE, ATT&CK).
14	Granata, Rak & Salzillo, 2022, [51]	Threat modeling	Estudo complementar (OSS)	Modelos e catálogos STRIDE/ATT&CK	—	Evidencia lacunas de cobertura, mas sem análise de desempenho de modelos automatizados.
15	Casola et al., 2020, [52]	Security-by-Design	Metodologia quantitativa via SLAs	Requisitos/SLA, métricas de serviço	—	Liga requisitos e métricas contratuais, porém depende de ambientes controlados e não cobre riscos dinâmicos.
16	Guggenmos et al., 2022, [33]	Estratégia em projetos	Security First / By Design / Pragmatism	Evidência organizacional	—	Destaca a importância cultural da segurança, mas sem métricas objetivas de impacto.
17	Lange & Kunz, 2024, [53]	Secure SDLC	Evolução de maturidade em soluções alojadas	Processos SDLC, controlos	—	Propõe framework evolutivo, mas aplicação depende da adesão organizacional e recursos.
18	Siavvas et al., 2023, [54]	SDLC / indústria	Monitorização de segurança no desenvolvimento	Métricas e eventos do ciclo	Métricas operacionais	Introduz medição contínua, mas enfrenta limitações na padronização entre pipelines distintos.

19	Akilal & Kechadi, 2022, [40]	Cloud forensics	Forensic-by-design (compliance)	Requisitos forenses/segurança	—	Traz rastreabilidade à cloud, mas pouca adoção prática e falta de automação de provas.
20	Buttar et al., 2023, [55]	Cloud / DevSecOps	Otimização DevOps/DevSecOps	Indicadores de pipeline	—	Reduz risco operacional, mas exige maturidade técnica e integração contínua entre equipas.
21	Patel et al., 2024, [56]	Threat Intelligence	Feeds de TI → deteção de exploração	Indicadores de ameaça, IOC/IOA	Taxa de deteção, F1	Antecipação de exploração, mas dependente da qualidade e atualidade dos feeds.
22	Torres et al., 2023, [24]	CTI regional (México)	Survey CTI orientado a dados	Feeds/relatos regionais	—	Identifica lacunas de dados e integração; limitada por contexto geográfico.
23	Zabihimayvan et al., 2024, [57]	Rede Tor	Survey sobre segurança/estrutura	Métricas e topologia Tor	—	Analisa anonimato e roteamento; limitação na coleta de dados reais de tráfego Tor.
24	Alotaibi et al., 2024, [58]	Arquiteturas BMC	Avaliação de resiliência a ataques	Componentes BMC, superfícies	—	Avalia superfícies de ataque em *Bare Machine Computing*, mas estudos ainda conceituais, sem validação empírica.

2.6.3. Lacunas Identificadas

Apesar dos avanços, a literatura evidencia lacunas relevantes que justificam investigações adicionais conforme compilado na Tabela 9.

Primeiramente, a maior parte dos trabalhos concentra-se em mecanismos de deteção de vulnerabilidades e ataques em tempo real — como demonstrado por Kumar et al. (2024) e Zhang et al. (2023), os quais fornecem respostas imediatas a incidentes, mas raramente abordam de forma explícita a previsão de riscos de cibersegurança em projetos tecnológicos. Esta ausência é significativa, pois, como salientam Guggenmos et al. (2022) e Noor et al. (2023), os projetos expandem continuamente a superfície de ataque ao longo do seu ciclo de vida, com novos ativos, integrações e dependências, exigindo uma abordagem preditiva capaz de antecipar riscos antes da sua materialização.

Em segundo lugar, a eficácia dos algoritmos depende fortemente da qualidade dos *datasets* utilizados. Guo et al. (2024) e Abdullah et al. (2022) observam que muitos estudos trabalham com dados desbalanceados, limitados a contextos específicos ou sem processos adequados de normalização e enriquecimento, o que, conforme destacam Dacorogna et al. (2023) e Haghighi & Ashrafi (2024), compromete a generalização dos modelos e a criação de soluções aplicáveis a diferentes ambientes organizacionais.

Por fim, permanece um desafio estrutural relacionado com a explicabilidade dos modelos (Explainable AI – XAI). Scholz et al. (2024) sublinham que técnicas como Deep Learning e Federated Learning oferecem elevada acurácia, mas operam como "caixas-pretas", sem mecanismos de justificação transparentes. De modo complementar, Patel et al. (2024) defendem que, em ambientes corporativos, a ausência de interpretabilidade reduz a confiança dos *stakeholders* e limita a adoção prática destas soluções, ainda que abordagens explicativas estejam a emergir de forma incipiente.

2.7. Conclusão da Revisão de Literatura

A revisão da literatura confirma que a Inteligência Artificial (IA) e o *Machine Learning* (ML) representam caminhos viáveis para reforçar a gestão de riscos em cibersegurança. Os estudos analisados demonstram que algoritmos clássicos, como *Decision Tree*, *Random Forest* e SVM, permanecem eficazes na classificação e priorização de vulnerabilidades [25] [45]. Trabalhos mais recentes evidenciam o potencial de redes neuronais profundas, técnicas híbridas de análise estática e dinâmica, bem como de abordagens ensemble (XGBoost, LightGBM) para lidar com ambientes complexos e *datasets* desbalanceados [46][36].

Além disso, observa-se uma evolução para arquiteturas mais avançadas, como *federated learning* aplicado a redes heterogéneas [48] e modelos *context-aware* para deteção multiestágio em sistemas ciberfísicos [37][38]. A literatura também destaca tendências emergentes, como o uso de Graph Neural Networks (GNNs) para representar relações entre ativos e ameaças [56] e a exploração de *Large Language Models* (LLMs) para extração de requisitos e análise de ameaças [59].

Apesar dessas contribuições, a literatura continua fragmentada: a maioria dos estudos foca-se em deteção reativa, negligenciando a previsão integrada de riscos de cibersegurança em projetos tecnológicos [23]. Persistem ainda os desafios relacionados à qualidade e diversidade dos *datasets* e à explicabilidade dos modelos, que comprometem a generalização e a confiança organizacional [18][59].

Assim, esta revisão fundamenta a relevância da presente dissertação: propõe um protótipo de modelo preditivo baseado em IA e ML, que integra *frameworks* normativos de riscos de cibersegurança, e ofereça um suporte dinâmico à tomada de decisão em projetos tecnológicos.

CAPÍTULO 3

Metodologia

3.1. Design Science Research

Esta dissertação segue a metodologia *Design Science Research* (DSR), numa implementação simplificada de seis etapas, conforma a Figura 4. Reconhece-se que uma abordagem completa de DSR incluiria múltiplos ciclos de refinamento, especialmente entre demonstração e avaliação. O presente trabalho implementa um único ciclo.

A escolha do método DSR, justifica-se pela natureza do problema abordado: a criação de um artefacto tecnológico baseado em IA para avaliação de riscos de cibersegurança, pois, oferece uma abordagem estruturada que permite não apenas o desenvolvimento iterativo do artefacto, mas também a sua validação contínua, para mostrar, a utilidade e a eficácia da solução proposta no contexto organizacional.

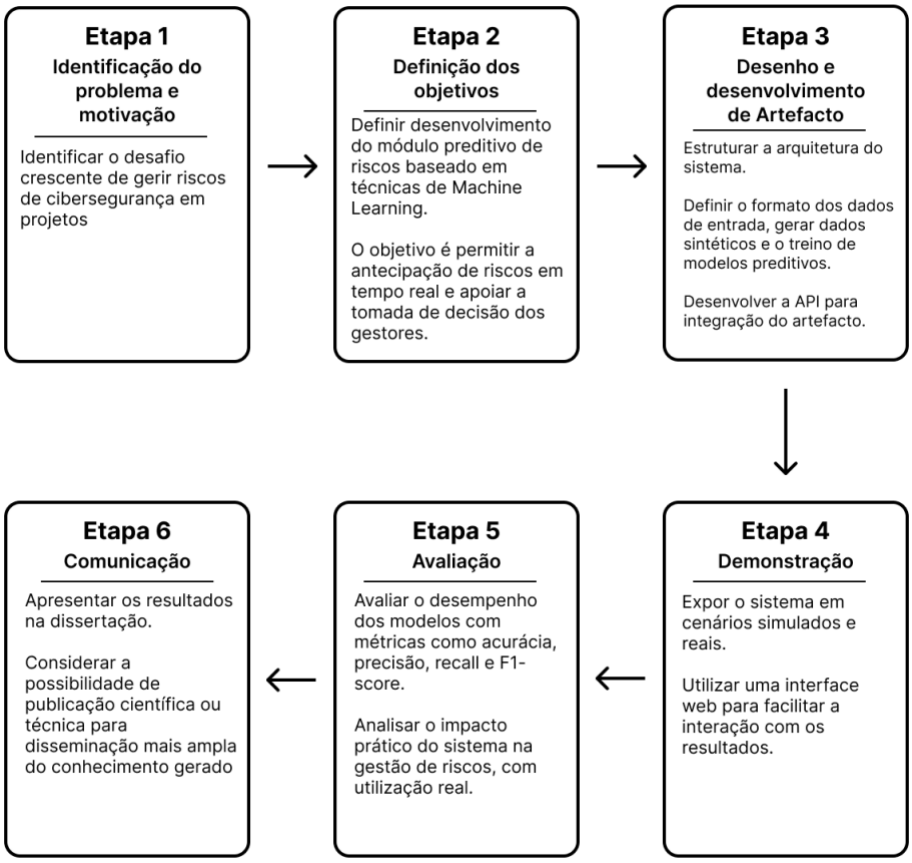


Figura 4 - Metodologia DSR

3.1.1. Identificação do Problema e Motivação

Identifica-se que a gestão proativa de riscos de cibersegurança é crucial no decurso de um projeto tecnológico. Pretende-se apresentar uma proposta de um modelo preditivo para gestão de riscos de cibersegurança que seja uma ferramenta capaz de ajudar os Gestores de Projetos e equipas de segurança na identificação antecipada de riscos, permitindo tratamento adequado antes da materialização de ameaças.

A principal motivação é contribuir com uma solução que permitirá monitorização contínua e preditiva de riscos de cibersegurança nos projetos, possibilitando a antecipação de potenciais problemas de segurança antes que se materializem. Isso oferecerá suporte valioso à tomada de decisão, transformando dados extraídos em insights acionáveis para cibersegurança.

3.1.2. Definição dos Objetivos para a Solução

O objectivo deste trabalho é apresentar um artefacto baseado em IA para auxiliar na gestão de riscos de cibersegurança. O artefacto deve processar dados estruturados provenientes de módulo adicional com dados sobre projetos, ativos, vulnerabilidades, ameaças que representam riscos e com algoritmos de ML e gerar previsão que apoiem decisões de mitigação de riscos de cibersegurança.

3.1.3. Design e Desenvolvimento do Artefacto

O desenvolvimento basear-se-á na utilização de algoritmos de ML aplicados á detenção e analise de riscos de cibersegurança, seguindo os pontos de identificados abaixo e alinhando-se com as melhores praticas descritas na literatura:

- Modelo Teórico: Arquitetura completa tem como base o processo processual para obter melhores resultado de ML nos riscos de cibersegurança;
- Protótipo com dados simulados replicando estrutura de dados para prova de conceito;
- Pré-processamento especializado para dados de cibersegurança;
- Treino modelos com algoritmos supervisionados de ML com dados de cibersegurança;
- Avaliação de resultados seguindo métricas específicas;
- Interface para visualização de previsão e recomendações de mitigação de riscos de cibersegurança.

3.1.4. Demonstração

A demonstração da utilidade do protótipo desenvolvido, dar-se-á com a utilização de dados simulados que replicam a estrutura dos dados que na prática possam vir de um sistema adicional, cobrindo diferentes tipos de projetos tecnológicos para demonstrar aplicabilidade em cenários variados de cibersegurança.

3.1.5. Avaliação

Na etapa de avaliação foram estabelecidas métricas específicas para modelos de ML, métricas como *precision*, *recall*, *F2-score*, *accuracy*, falsos positivos e negativos. Adicionalmente realização de testes de usabilidade e experiência com profissionais reais, com recolha de feedbacks no formato de inquérito.

3.1.6. Comunicação dos Resultados

Documentar o desenvolvimento, as demonstrações e os resultados. Pela elaboração da dissertação, igualmente, publicação do resultado em artigo científico.

CAPÍTULO 4

Desenvolvimento do Modelo Preditivo

4.1. Definição da Estrutura de Dados para Modelagem

Conforme verificado na revisão da literatura, a qualidade e a organização dos dados são fatores essenciais em ML e para o desenvolvimento de modelos preditivos eficazes.

A estrutura de dados criada foi pensada para ser flexível e adaptável a diferentes tipos de projetos tecnológicos, permitindo que o modelo preditivo possa receber e analisar informações provenientes de múltiplos contextos.

Nas seções seguintes, são apresentados os detalhes da construção dessa estrutura, bem como as fontes científicas e bases de referência que fundamentaram a definição de cada bloco e variável do *dataset*.

Como etapa inicial, estabeleceu-se que a principal fonte para estruturar os dados são base de recomendações científica, incluindo informações sobre projetos e riscos identificados, aponta PMBOK e ISO 31000 conforme a Tabela 10.

Tabela 10 - Features recomendadas por PMBOK e ISO 31000 no registo de risco

Feature	Descrição	Reference
project_type	Tipo de projeto associado ao risco.	PMBOK
risk_type	Natureza do risco identificado (ex: atraso no cronograma, indisponibilidade de recursos).	PMBOK / ISO 31000
risk_category	Classificação do risco em áreas (tempo, escopo, recursos, etc.).	PMBOK
probability	Probabilidade qualitativa de ocorrência (ex: "Rare", "Possible", "Frequent").	ISO 31000 / PMBOK
Impact	Nível qualitativo de impacto (ex: "Minor", "High", "Critical").	ISO 31000 / PMBOK
severity	Combinação de impacto e probabilidade (matriz de risco).	ISO 31000 / PMBOK
strategy	Estratégia de resposta ao risco (ex: Mitigar, Transferir, Aceitar).	PMBOK
timestamp	Data/hora do registro do risco.	ISO 31000 (como boa prática)

PMBOK (7ª Edição) - define a estrutura do registo de risco no contexto de gerenciamento de projetos, recomendando campos como tipo, categoria, e resposta ao risco a estratégia [15].

ISO 31000:2018 - fornece diretrizes para gestão de riscos em qualquer organização, com foco em probabilidade, impacto, severidade e processo contínuo de avaliação [13].

No entanto, para a flexibilidade e atender os objectivo delineados, reconstrui-se e definiu-se o *dataset* para cobrir aspectos de cibersegurança.

A definição da estrutura de dados representa um passo crítico na construção do artefacto preditivo. Embora o PMBOK (7.ª edição) e a ISO 31000:2018 forneçam diretrizes sólidas sobre o registo de riscos, tais elementos são insuficientes quando o foco é a cibersegurança. Para este domínio, torna-se necessário estender o modelo de dados de forma a capturar três dimensões fundamentais: ativos, vulnerabilidades e ameaças.

As *features* que compõem o *dataset* deste estudo foi orientada por princípios reconhecidos de gestão e análise de riscos de cibersegurança, bem como por *frameworks* normativos internacionais (NIST, ISO, ENISA, CIS, MITRE, OWASP), conforme a Tabela 11.

Tabela 11 – Fontes e base da geração dos dados

Tipo	Fontes	Observação
Vulnerabilidades	NIST - CyberSecurity Framework, CVE, NVD, CWE [11]	*Features* sobre, Vulnerabilidades tipo de riscos e severidade
Ameaças	MITRE ATT&CK [43], CISA KEV, Malpedia	Mapear táticas e ataques reais
Ativos	CPE, Shodan	Associar vulnerabilidades a sistemas específicos
Riscos	VERIS, CIS Controls	Criar rótulos (impacto, probabilidade, severidade)

Organizadas em blocos de governança, ativos, vulnerabilidades, ameaças e risco, baseada nas melhores praticas [19][21], com a respetiva tipologia de dados conforme apresentado na Tabela 12.

Tabela 12 - Features de Cibersegurança

Bloco	Feature	Descrição	Tipo
Projeto	project_type	Tipo de projeto (ex: software, blockchain, web_app)	Categórica
Risco	risk_type	Tipo de risco identificado (ex: ddos_disruption, zero_day_exploit)	Categórica
	risk_category	Categoria do risco (ex: stakeholder, compliance)	Categórica
	risk_confirmed	Indicador se o risco se confirmou (0 = Não, 1 = Sim)	Binária
Gravidade	severity_score	Escore de severidade do risco	Numérica
	Cvss	Common Vulnerability Scoring System (0–10)	Numérica
	Epss	Exploit Prediction Scoring System	Numérica
Probabilidade	Probability	Probabilidade calculada do risco ocorrer	Numérica
Ativo	asset_type	Tipo de ativo sendo avaliado (ex: database, web_server)	Categórica
Segurança do Ativo	internet_facing	Se o ativo está exposto à internet (0 = Não, 1 = Sim)	Binária
	business_critical	Se o ativo é crítico para o negócio (0 = Não, 1 = Sim)	Binária
	Mfa	Presença de autenticação multifator (0 = Não, 1 = Sim)	Binária
Defesa	Edr	Endpoint Detection & Response habilitado (0 = Não, 1 = Sim)	Binária
Resiliência	backup_immutable	Backups imutáveis disponíveis (0 = Não, 1 = Sim)	Binária
Controles	control_coverage_score	Cobertura de controles aplicada ao ativo	Numérica
Vulnerabilidade	Vulnerability	Vulnerabilidade relacionada ao ativo (ex: patch_noncompliance)	Categórica
	vuln_age_days	Idade da vulnerabilidade em dias	Numérica
Patch Management	patch_sla_adherence	Conformidade com SLA de patches	Numérica (%)
Ameaça	Threat	Tipo de ameaça cibernética detectada (ex: malware_delivery)	Categórica
Inteligência Ameaça	is_kev	Se a vulnerabilidade é conhecida em KEV (0 = Não, 1 = Sim)	Binária
Monitoramento	alerts_30d	Quantidade de alertas nos últimos 30 dias	Numérica
Ataques	web_attacks_30d	Número de ataques web detectados em 30 dias	Numérica
	malware_detected_30d	Número de malwares detectados nos últimos 30 dias	Numérica
	bruteforce_30d	Tentativas de brute force nos últimos 30 dias	Numérica
Temporalidade	snapshot_at	Data de geração da instância	Temporal
	label_window_days	Janela de tempo usada para rotular o risco	Numérica
	gap_days	Dias de diferença entre evento e janela de análise	Numérica

Sobre os riscos abaixo na Tabela 13 apresenta-se uma lista de possíveis tipos de riscos no contexto de cibersegurança.

Tabela 13 - Tipos de Riscos de cibersegurança

Risk Type	Descrição
security_breach	Acesso não autorizado a sistemas, dados ou redes, incluindo ataques internos ou externos.
data_loss	Perda ou corrupção de dados críticos por falhas técnicas, malware, ou erro humano.
ransomware_outage	Indisponibilidade prolongada devido a ataques de ransomware.
ddos_disruption	Interrupção de serviços causada por ataques de negação de serviço distribuído (DDoS).
zero_day_exploit	Exploração de vulnerabilidades não documentadas (zero-day).
supply_chain_compromise	Comprometimento de fornecedores externos ou dependências de software [60].
credential_stuffing	Uso de credenciais roubadas para acesso a sistemas críticos.
phishing_takeover	Aquisição indevida de contas através de ataques de phishing.
api_abuse	Exploração indevida de APIs, levando a fuga de dados ou interrupção de serviços.
misconfiguration_exposure	Exposição indevida de sistemas devido a configurações incorretas.
insider_misuse	Abuso intencional ou negligência por colaboradores com acesso privilegiado.
vulnerability_backlog	Acumulação de vulnerabilidades críticas não corrigidas no tempo adequado.
patch_noncompliance	Atrasos ou falhas recorrentes em aplicar correções de segurança.
integration_issue	Problemas ao integrar novos sistemas, gerando falhas de segurança.
regulatory_noncompliance	Falhas em atender normas legais e regulatórias (ex.: GDPR, ISO 27005).

Os riscos também recebem uma categorização, assim, abaixo apresenta-se uma lista das possíveis categorizações na Tabela 14.

Tabela 14 - Categorização dos Riscos voltado a cibersegurança

Risk Category	Descrição
Security	Riscos ligados a ataques, violações de dados e incidentes de confidencialidade, integridade e disponibilidade.
Technical	Riscos de falhas tecnológicas, bugs, vulnerabilidades e arquiteturas complexas.
Operational	Riscos no funcionamento contínuo dos sistemas, como *downtime* e falhas de processos.
Compliance	Riscos de não conformidade com leis, regulamentos e normas de segurança.
Financial	Perdas monetárias diretas ou indiretas decorrentes de incidentes de cibersegurança.
Schedule	Atrasos em projetos tecnológicos por incidentes ou exigências de mitigação de riscos.
Resource	Limitação de pessoal qualificado, ferramentas de segurança ou infraestrutura.
Stakeholder	Conflitos ou perda de confiança de clientes, reguladores e parceiros devido a falhas de segurança.
Contractual	Problemas relacionados a acordos com fornecedores e parceiros que impactam segurança.
Integration	Riscos de segurança na integração de novos sistemas, plataformas ou APIs.
Human	Riscos associados ao fator humano: erro, negligência ou ataques internos.
Design	Riscos devido a falhas ou omissões na conceção segura de sistemas.
Data	Riscos ligados à perda, roubo, manipulação ou baixa qualidade de dados.

Com as estruturas acima definidas, a base dos dados sobre projetos, riscos de cibersegurança estão estabelecidas.

4.2. Arquitetura do Modelo Preditivo de Risco de Cibersegurança

Com os objectivos direcionado para garantir um protótipo de previsão de riscos de cibersegurança apresenta-se um fluxo conceptual conforme a Figura 5.

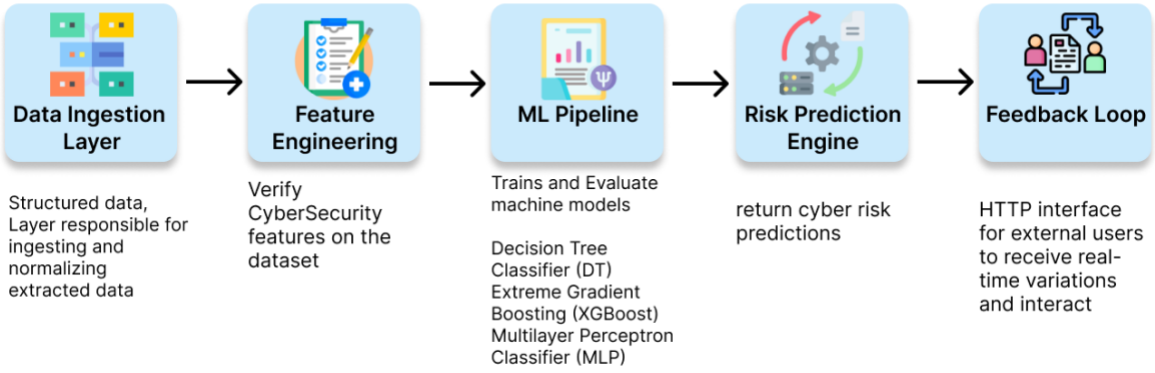


Figura 5 – Fluxo de processamento do protótipo

Destaca que o protótipo tem na base recepção de dados estruturados, aplica *feature engineering* sobre os dados e realiza o processo de treino com os algoritmos de ML para previsão de riscos de cibersegurança.

Num contexto em produção e abrangente o fluxo conceptual integra módulos adicionais, como o NLP para extração de dados em diferentes tipo de documentos afeto aos projetos, e isso ficou espelhado na arquitetura do modelo preditivo conforme a Figura 6.

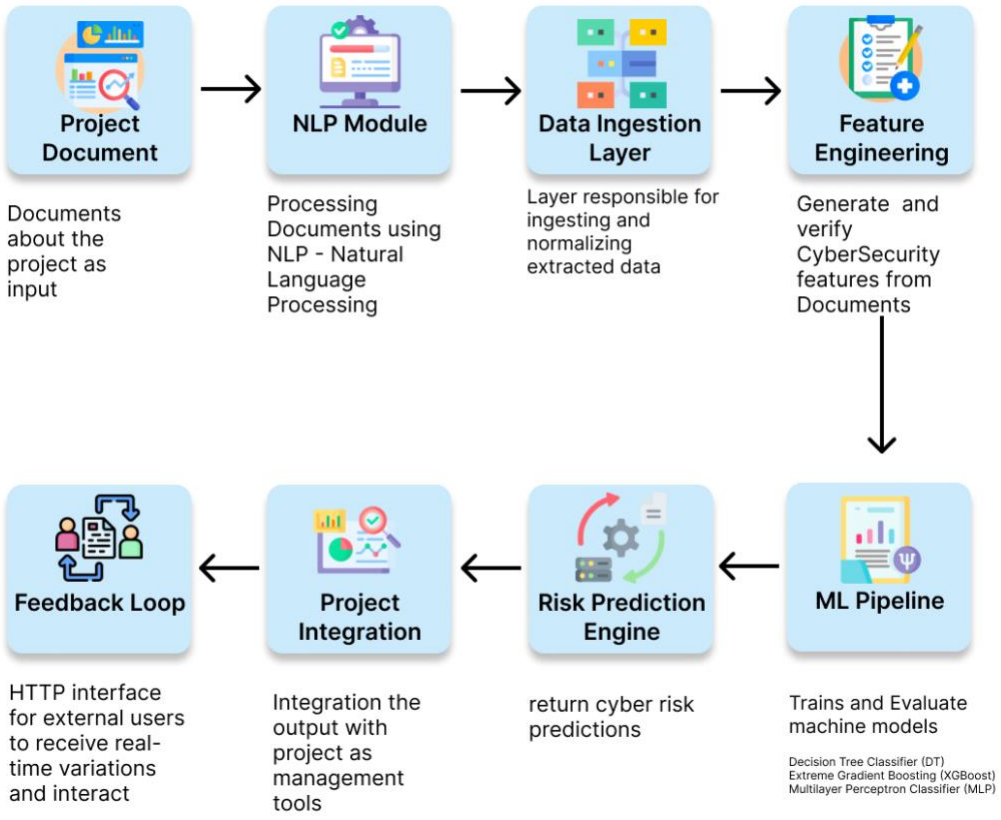


Figura 6 - Fluxo de Completo de processamento

Cada médulo do fluxo conceptual é descrito abaixo:

- **Project Document**: Corresponde aos documentos e registos do projeto que servem de input inicial. Contêm informações relevantes sobre requisitos, cronogramas, recursos e dependências, que podem revelar potenciais riscos de cibersegurança.
- **NLP Module**: Módulo de *Natural Language Processing* (NLP) responsável por processar automaticamente os documentos do projeto, extraindo entidades e termos-chave relacionados com ameaças, vulnerabilidades e controlos de segurança.
- **Data Ingestion Layer**: Camada responsável por ingerir, normalizar e armazenar os dados extraídos. Esta fase garante a consistência e qualidade dos dados antes de entrarem na etapa analítica, preparando-os para posterior transformação em *features*.
- **Feature Engineering:** Etapa dedicada à geração de variáveis (características) relevantes para o modelo de previsão de riscos. A partir dos dados extraídos, são criados indicadores que representam padrões de exposição, vulnerabilidade e impacto potencial.
- **ML Pipeline**: Conjunto de procedimentos automáticos que treina e avalia algoritmos de Machine Learning (como *Decision Tree Classifier*, *Extreme Gradient Boosting* e *Multilayer Perceptron*). Esta camada produz modelos preditivos capazes de estimar a probabilidade e severidade de riscos.
- **Risk Prediction Engine:** Núcleo de inferência do sistema. Utiliza o modelo treinado para gerar previsões de risco, classificando os resultados conforme a probabilidade de ocorrência e impacto em cada ativo ou componente do projeto.
- **Project Integration**: Responsável pela integração dos resultados com ferramentas de gestão de projetos e plataformas externas. Permite que as previsões de risco sejam visualizadas e interpretadas no contexto operacional do projeto.
- **Feedback Loop**: Mecanismo que recolhe dados de retorno e eventos reais do ambiente (por exemplo, incidentes ocorridos ou riscos confirmados). Esta retroalimentação permite ajustar o modelo, melhorando continuamente a precisão e adaptabilidade das previsões.

4.3. Pré-processamento dos Dados

Após uma visão do fluxo de processamento do protótipo, aqui referimos a atenção no processamento dos dados estruturado em *dataset*, apesar de ter com base a estrutura conforme a Tabela 12, garante-se o processo de pré-processamento com atenção em *features* que sustentam normas, e frameworks sobre gestão de riscos de cibersegurança.

Validação de *schema* dos dados:

- Normalização dos impacto e probabilidade segundo o PMBOK e escala de severidade;
- Codificação de variáveis categóricas através de *StringIndexer* e *OneHotEncoder*;
- Engenharia de características temporais e correlacionais;
- Tratamento de dados desbalanceados e validação de nulos e tipos.

4.4. Treinamento e Avaliação dos Modelos

Nesta investigação, selecionou-se três algoritmos supervisionados para o problema de riscos em cibersegurança nos projetos tecnológicos: *Extreme Gradient Boosting* (XGBoost), *Multilayer Perceptron* (MLP) e *Decision Tree* (DT). A escolha justifica-se pela combinação entre robustez em cenários de dados desbalanceados, capacidade de modelar relações complexas e níveis distintos de interpretabilidade, fatores críticos na previsão de riscos de cibersegurança [17], [19], [38], [50]. Por inferência do que o estado da arte, a revisão da literatura mostrou nos estudos sobre ML aplicada a cibersegurança.

4.4.1. Extreme Gradient Boosting (XGBoost)

O XGBoost é um algoritmo de *ensemble learning* baseado em árvores de decisão que implementa *gradient boosting*, otimizando uma função de perda através de adições sequenciais de árvores fracas. Formalmente, o modelo aprende funções aditivas $f_t(x)$ de forma a minimizar [61]:

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y_i}) + \sum_{i=1}^{T} \Omega(f_t)$$

Onde l é a função de perda (por exemplo, *log loss* para classificação binária) e Ω é o termo de regularização que controla a complexidade. Introduzido por Chen & Guestrin (2016) [61], rapidamente se tornou referência. Aplicações em cibersegurança: amplamente usado em deteção de intrusões, classificação de vulnerabilidades e análise de malware devido à sua performance em *datasets* desbalanceados [19], [38], [61]. Possui os seguintes parâmetros relevantes: *learning rate*, *max_depth*, *n_estimators*, *subsample*. Tem se mostrado robusto contra ruído, escalável e fornece interpretabilidade via *feature importance*.

4.4.2. Multilayer Perceptron (MLP)

O MLP é uma rede neural *feedforward* composta por camadas densas, com funções de ativação não lineares (ReLU, *sigmoid*, tanh). O treino é realizado via retropropagação (*backpropagation*), minimizando a função de perda L com gradiente descendente:

$$\Delta w_{ij} = -\eta \frac{\partial L}{\partial w_{ij}}$$

Onde w_{ij} são os pesos e η é a taxa de aprendizagem. Concebido nos anos 80, foi revalorizado com avanços em hardware e técnicas de regularização. Eficaz em deteção de padrões complexos em tráfego de rede, análise de anomalias e correlação de eventos [17], [19], [26]. Parâmetros relevantes: número de camadas ocultas, neurónios por camada, função de ativação, *dropout*. Permite capturar relações não-lineares entre variáveis de risco, sendo adequado quando a complexidade dos dados vai além da capacidade de modelos baseados em árvores.

4.4.3. Decision Tree (DT)

As árvores de decisão criam um modelo hierárquico, segmentando o espaço de atributos por regras do tipo if-then. A impureza de cada nó pode ser medida, por exemplo, via índice Gini:

$$Gini = 1 - \sum_{i=1}^{C} p_i^2$$

Onde p_i é a proporção da classe i no nó. Popularizadas por Breiman et al. (1984) [62], no método CART (*Classification and Regression Trees*) [64]. usadas em classificação de *malware*, deteção de intrusões e análise de vulnerabilidades [19], [65]. Parâmetros relevantes: *max_depth*, *min_samples_split*, *criterion* (Gini, Entropy). É um dos modelos mais interpretáveis, o que favorece a explicabilidade em contextos de auditoria e tomada de decisão, embora apresente riscos de *overfitting*.

4.4.4. Métricas para Avaliação dos modelos ML

A avaliação dos modelos de ML em cibersegurança recorre a métricas de classificação amplamente reconhecidas, cada uma refletindo um aspeto distinto do desempenho do classificador. A *accuracy* (exatidão) mede a proporção de previsões corretas em relação ao total de instâncias avaliadas, sendo expressa pela fórmula:

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

Em que TP representa os verdadeiros positivos, TN os verdadeiros negativos, FP os falsos Positivos e FN os falsos negativos [66]. Apesar da sua simplicidade e utilidade como métrica geral, a *accuracy* pode ser enganadora em *datasets* desbalanceados, típicos da cibersegurança [66].

A *precision* (precisão) mede a proporção de previsões positivas que são de facto corretas, sendo calculada como:

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

Esta métrica é particularmente relevante em cibersegurança, uma vez que contribui para reduzir falsos positivos, evitando a sobrecarga de alarmes indevidos em sistemas de deteção de intrusões [19]. Complementarmente, o *recall* (sensibilidade ou revocação) avalia a capacidade do modelo em identificar corretamente todas as instâncias positivas, definido pela expressão:

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

Em contextos de cibersegurança, o *recall* assume importância crítica, pois ataques não detetados (falsos negativos) podem comprometer seriamente a integridade organizacional [17].

Com o intuito de equilibrar os *trade-offs* entre *precision* e *recall*, adota-se o F1-score, que corresponde à média harmónica entre estas duas métricas:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Este indicador mostra-se particularmente útil em cenários com classes desbalanceadas, como é comum em *datasets* de tráfego de rede, onde registos maliciosos representam uma minoria [5].

Por fim, a AUC-ROC (*Area Under the Curve - Receiver Operating Characteristic*) avalia a capacidade discriminatória de um modelo em diferentes limiares de decisão [62]. Formalmente, pode ser expressa como a integral da taxa de verdadeiros positivos (TPR) em função da taxa de falsos positivos (FPR):

$$AUC = \int_0^1 TPR(FPR)d(FPR)$$

sendo $TPR = \frac{TP}{TP+FN}$ e $FPR = \frac{FP}{FP+TN}$. Valores próximos de 1 indicam elevada capacidade do classificador em distinguir instâncias positivas de negativas, o que é desejável em sistemas de deteção de intrusões e análise de vulnerabilidades [18].

Deste modo, cada métrica acrescenta uma perspetiva específica sobre o desempenho dos algoritmos testados: enquanto a *accuracy* fornece uma visão global, a *precision* e o *recall* capturam dimensões complementares de eficácia, o F1-score equilibra estas duas vertentes e a AUC-ROC avalia a robustez da classificação em múltiplos cenários de decisão.

CAPÍTULO 5

Implementação e Validação do Protótipo

5.1. Ambiente de Desenvolvimento

O protótipo foi desenvolvido utilizando na sua base o Python 3.9 como linguagem principal, aproveitando seu ecossistema robusto para ML e processamento de dados. O ambiente de desenvolvimento incluiu:

Bibliotecas Principais:

- scikit-learn (1.3.0): Para algoritmos de ML e métricas de avaliação
- pandas (2.0.3): Manipulação e análise de dados tabulares
- numpy (1.24.3): Computação numérica e operações matriciais
- matplotlib/seaborn: Visualização de dados e resultados
- XGBoost (1.7.0): Implementação otimizada do gradient boosting
- FastAPI (0.115.13): Framework para APIs REST

Infraestrutura:

- Sistema operacional: macOS Sequoia 15.6.1
- Hardware: 2 GHz Quad-Core Intel Core i5, 16GB RAM, SSD 1TB
- Containerização: Docker para isolamento de dependências no PostgresSQL
- Versionamento: Git com repositório privado: https://github.com/fjpiedade/risk-manager

Utilização do Python justifica-se por ser uma linguagem amplamente utilizada em ML pela simplicidade, flexibilidade e grande ecossistema de bibliotecas, permitindo desde o tratamento de dados até a implementação de modelos preditivos em produção.

5.2. Diagrama do Protótipo Preditiva de Riscos de Cibersegurança

Implementou-se o modelo preditivo para riscos de cibersegurança com base na arquitetura, conforme se verifica na Figura 7, estruturado em cinco componentes principais dentro do fluxo conceptual abordado no capítulo anterior na Figura 5:

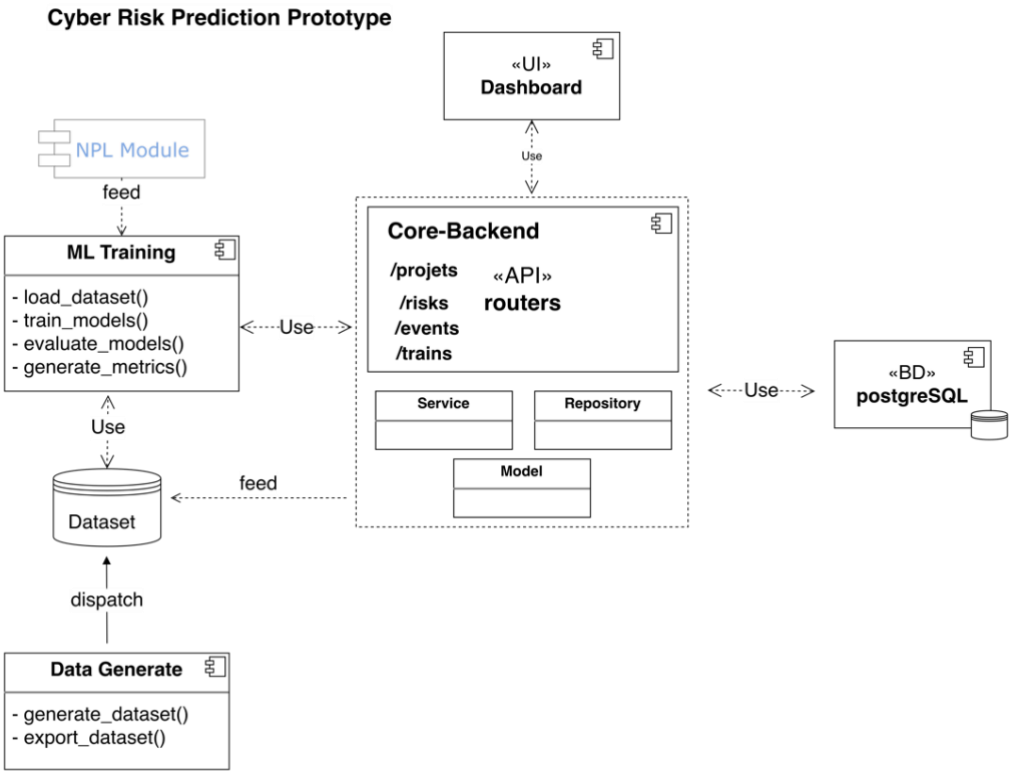


Figura 7 – Diagrama de componentes do protótipo de previsão de riscos

Data Generate

Responsável pela criação e exportação dos *datasets* sintéticos (generate_dataset(), export_dataset()), usados para treino e avaliação dos modelos. Simula dados de projetos, eventos e riscos com base em cenários realistas.

ML Training

Módulo de treino de ML que utiliza os dados gerados. Executa funções como carregamento de dados (load_dataset()), treino dos modelos (train_models()), avaliação (evaluate_models()) e geração de métricas (generate_metrics()).

O resultado é um conjunto de modelos preditivos capazes de identificar riscos com base nos padrões aprendidos.

API Routers

Núcleo central do sistema, implementado em FastAPI. Contém os endpoints principais (/projects, /risks, /events, /trains), organizados por camadas — Model, Repository e Service — para garantir modularidade e separação de responsabilidades.

Aos detalhes, o sistema possui os seguintes *endpoints*:

`POST /projects` – Cria um novo projeto, regista seus ativos e executa a predição de riscos com modelos de ML. O retorno inclui o identificador do projeto, a lista de riscos previstos, a severidade agregada e os ativos, nessa requisição tem como obrigatoriedade informar o tipo de projeto, opcionalmente se pretende ter previsão sobre ativos específicos, e escolha do algoritmo que pretende usar para previsão dos riscos cibernéticos, conforme se verifica na Figura 8.

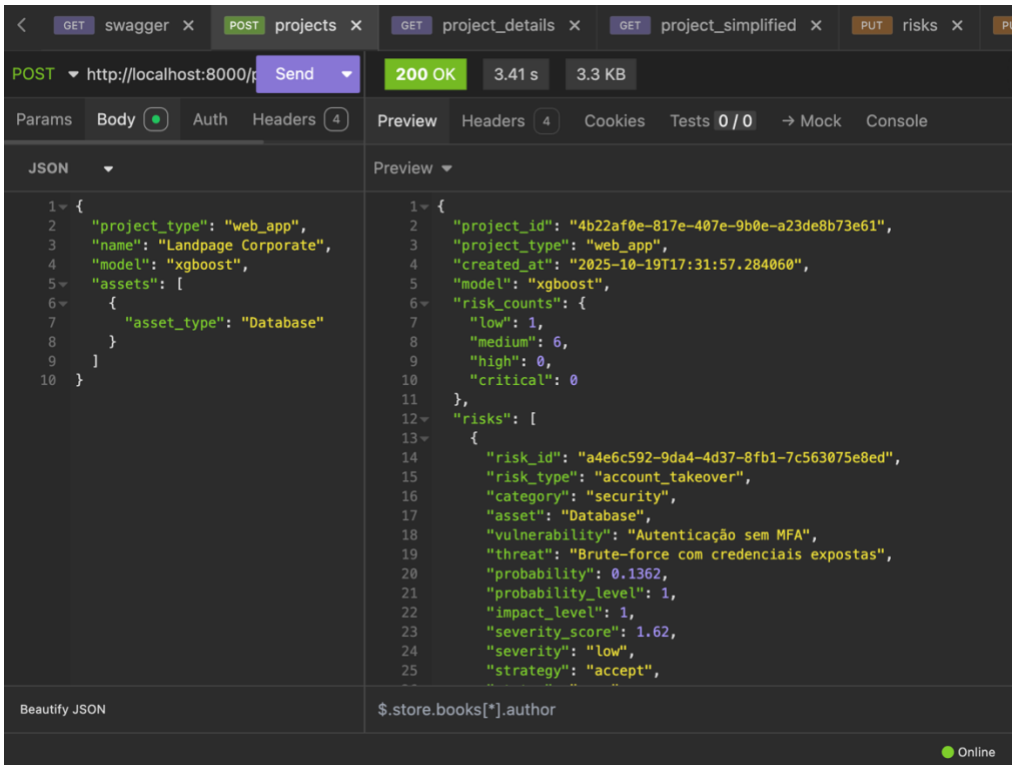


Figura 8 - Requisição para previsão de riscos

`GET /projects` – Lista todos os projetos cadastrados no sistema, retornando para cada um o identificador, nome, tipo, data de criação e a contagem de riscos por nível de severidade.

`GET /projects/{project_id}` – Consulta detalhadamente um projeto específico, incluindo os ativos associados e os riscos detetados.

`PUT /risks/{risk_id}` – Atualiza um risco existente, criando uma nova versão quando há alterações relevantes. Permite também registar feedback humano (ação tomada, comentários e confirmação). Caso o risco seja fechado e confirmado, ele é automaticamente inserido no dataset de treinamento para futura atualização do modelo.

`GET /risks/{risk_id}/history` – Recupera o histórico de versões de um risco, exibindo todas as mudanças realizadas, feedbacks registados e evolução do status ao longo do tempo.

`POST /models` – Executa o processo de treinamento dos modelos de *Machine Learning* com base no *dataset* atualizado, incorporando novos riscos confirmados.

`POST /projects/{project_id}/events` – Regista um evento em um projeto e aplica sua lógica de impacto nos riscos. Se o evento estiver mapeado, ele pode mitigar, aumentar ou fechar riscos já existentes. Caso contrário, é realizada inferência por ML para sugerir novos riscos. O retorno contém o identificador do evento, seu impacto e os riscos afetados.

`GET /projects/{project_id}/events` – Lista todos os eventos registados em um projeto, retornando identificadores, tipos, descrições, impactos e datas de ocorrência.

Com essa arquitetura outros sistemas de gestão de projetos facilmente se podem integrar em *realtime* e ter uma visão dos riscos de cibersegurança ao longo do ciclo de vida do projeto.

Esta API comunica com o banco de dados e serve de interface entre o motor preditivo e o *dashboard*.

PostgreSQL (BD)

Base de dados relacional utilizada para armazenar projetos, eventos, riscos e feedbacks do utilizadors, como de PM. Garante persistência, integridade e rastreabilidade das previsões e feedbacks recebidos.

UI Dashboard

Interface de utilizador que apresenta os resultados de forma visual e interativa. Permite monitorizar previsões de risco, e o estado dos projetos.

O diagrama inclui ainda um módulo de *Natural Language Processing* (NLP), previsto, mas não implementado na versão atual. Este componente permitirá processar automaticamente documentos de projeto, extraindo informação relevante para gerar *datasets* e enriquecer a camada de ingestão de dados, ampliando a capacidade preditiva de riscos.

5.3. Implementação da pipeline do Protótipo

A implementação do pipeline do protótipo é uma etapa central do protótipo, pois, permite organizar e executar a sequência lógica de atividades que vão desde a ingestão de dados até a análise dos resultados. Utilizou-se scripts, desenvolvido em Python, para processar os dados de entrada e aplicar os algoritmos de ML escolhidos como referenciado no Capítulo 4.

5.3.1. Geração de Dados Sintéticos

Os dados para sustentar os modelos sempre constituem um enorme peso no processo, nesse caso de estudo, os dados foram gerados tendo com base os repositórios, NIST - CyberSecurity Framework, MITRE ATT&CK Framework [43], a CVE Database (2020–2025) [11] e o OWASP Top 10. As Tabela 11, Tabela 13 e Tabela 14 apresentam a estrutura adotada e a relação entre as variáveis e suas fontes de referência.

Para garantir que os dados sintéticos estivessem em conformidade com normas e classificações reconhecidas, o processo de geração seguiu as categorias e escalas dessas fontes, buscando manter coerência e representatividade em relação a contextos reais de risco cibernético. Com isso, o gerador usado no protótipo é um script python, controlado localmente respeitando os pressupostos da Tabela 11.

O gerador dos dados é apresentado no Apêndice A. Este gerador simula contextos realistas de projetos tecnológicos e ataques de cibersegurança, permitindo a criação de conjuntos de dados diversos, com diferentes perfis de risco. Essa abordagem garante maior controle sobre as variáveis e facilita a experimentação, mesmo na ausência de grandes volumes de dados reais.

Dataset:

- 5000 entradas simuladas de riscos, conforme o Gráfico 2;
- tipos de projetos: "cloud", "onprem", "hybrid", "iot_system", "mobile_app", "web_app","blockchain_platform", "ai_model_deployment", "api_gateway";
- Distribuição realística de vulnerabilidades baseada em OWASP Top 10;
- Padrões de ameaças derivados de MITRE ATT&CK framework.

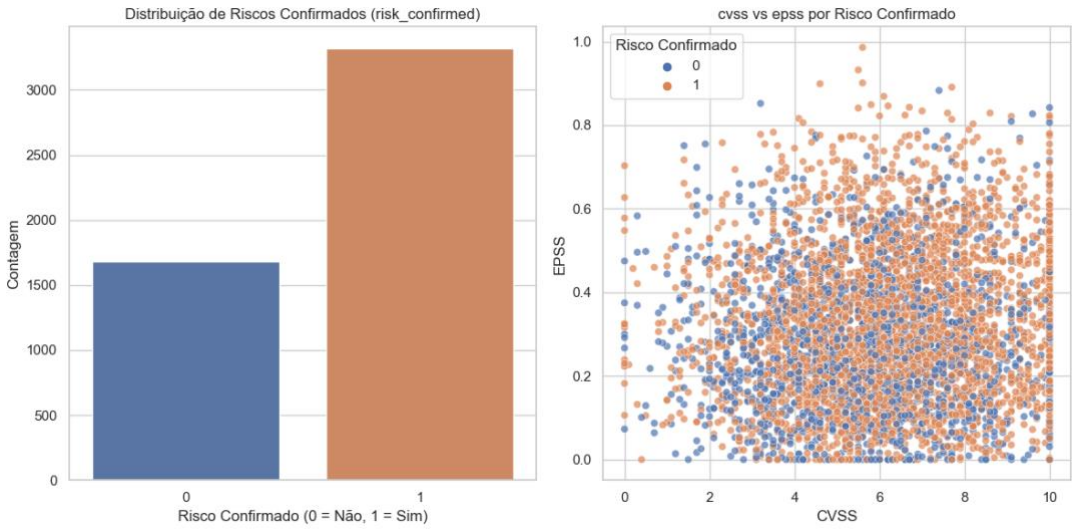


Gráfico 2 - Destruição dos dados utilizados

O Gráfico 2 mostra a distribuição dos dados, do *dataset* usado, 5000 entradas, ~70% com riscos confirmado e ~30% não confirmados e distribuição dos epss e cvss.

5.3.2. Implementação dos Modelos ML

A construção dos modelos de ML é realizada no *script*, onde são treinados e validados os três algoritmos selecionados, e há uma divisão dos dados conforme a lista abaixo. Cada modelo é preparado para ser utilizado diretamente pela aplicação, de forma modular e integrada.

Divisão dos Dados:

- Treino: 70% dos dados
- Validação: 15% dos dados
- Teste: 15% dos dados

Para cada tipo de risco (risk_type), o sistema treina e calibra três modelos distintos (XGBoost, MLP e DecisionTree).

Cada modelo treinado é exportado em dois ficheiros complementares:

- Modelo (.joblib): contém o objeto de Machine Learning serializado, pronto para inferência.
- Metadados (.json): documenta o desempenho e as condições do treino, incluindo métricas, data, threshold e caminho do ficheiro binário.

Estes ficheiros permitem a rastreabilidade total dos modelos, facilitando auditoria, reprodutibilidade e versionamento no processo de previsão de riscos.

Nesse mesmo processo conforme mostrou o diagrama na Figura 7, são gerado as métricas para cada algoritmos ML em relação aos dados.

5.3.3. Resultados dos Modelos ML

Na Tabela 15 abaixo pode se verificar as métricas obtidas para cada modelo e incidem sobre tipos de riscos de cibersegurança, realçando que foi com base nos dados sintéticos e com as configurações de parâmetros dos algoritmos de ML conforme o Apêndice F:

Tabela 15 - Métricas dos Modelos Treinados de ML

risk_type	model	accuracy	precision	recall	f1_score	f2_score	roc_auc	threshold	tp	fp	tn	fn
security_breach	XGBoost	0.68	0.67	0.99	0.80	0.91	0.62	0.6	165	80	4	1
security_breach	MLP	0.66	0.66	1.00	0.80	0.91	0.53	0.1	166	84	0	0
security_breach	DecisionTree	0.66	0.67	0.98	0.80	0.90	0.51	0.6	163	81	3	3
data_loss	XGBoost	0.69	0.73	0.88	0.80	0.84	0.60	0.6	162	60	25	23
data_loss	MLP	0.69	0.69	1.00	0.81	0.92	0.56	0.1	185	85	0	0
data_loss	DecisionTree	0.69	0.69	1.00	0.81	0.92	0.56	0.1	185	85	0	0
zero_day_exploit	XGBoost	0.61	0.61	1.00	0.76	0.89	0.57	0.1	158	99	0	0
zero_day_exploit	MLP	0.62	0.62	0.96	0.76	0.86	0.53	0.6	151	91	8	7
zero_day_exploit	DecisionTree	0.61	0.61	1.00	0.76	0.89	0.46	0.1	158	99	0	0
supply_chain_compromise	XGBoost	0.65	0.66	0.99	0.79	0.90	0.57	0.6	154	81	2	2
supply_chain_compromise	MLP	0.65	0.65	1.00	0.79	0.90	0.52	0.1	156	83	0	0
supply_chain_compromise	DecisionTree	0.66	0.66	0.99	0.79	0.90	0.54	0.6	154	80	3	2
ddos_disruption	XGBoost	0.70	0.70	1.00	0.82	0.92	0.55	0.1	163	71	0	0
ddos_disruption	MLP	0.70	0.70	1.00	0.82	0.92	0.52	0.1	163	71	0	0
ddos_disruption	DecisionTree	0.70	0.70	0.98	0.82	0.91	0.48	0.7	160	68	3	3
ransomware_outage	XGBoost	0.67	0.67	1.00	0.80	0.91	0.54	0.1	169	83	0	0
ransomware_outage	MLP	0.67	0.67	1.00	0.80	0.91	0.52	0.1	169	83	0	0
ransomware_outage	DecisionTree	0.67	0.67	1.00	0.80	0.91	0.55	0.1	169	83	0	0

A Tabela 15 apresenta as métricas de desempenho dos modelos XGBoost, Multilayer Perceptron (MLP) e Decision Tree (DT), aplicados aos diferentes tipos de risco de cibersegurança analisados (*ddos_disruption, security_breach, zero_day_exploit, supply_chain_compromise, ransomware_outage e data_loss*). Foram avaliadas as métricas de *accuracy, precision, recall, F1-score, F2-score* e *AUC-ROC*, bem como os valores de *threshold* e as contagens de verdadeiros e falsos positivos e negativos, (FP) e (FN). Esta análise visa compreender o comportamento dos modelos em contextos distintos e avaliar a sua capacidade de generalização em cenários potencialmente desbalanceados.

De forma geral, observa-se que nenhum modelo apresenta supremacia absoluta em todas as métricas e categorias de risco. Os três algoritmos exibem desempenhos próximos, com valores de F1-score entre 0.76 e 0.82, *recall* elevado (≥ 0.96) e *precision* moderada (≈ 0.65–0.70). Este comportamento indica que o conjunto de dados possui padrões suficientemente homogéneos para serem capturados por diferentes arquiteturas, resultando em níveis de desempenho semelhantes. As diferenças entre modelos concentram-se, sobretudo, no equilíbrio entre *precision* e *recall*, o que influencia diretamente o número de falsos positivos e falsos negativos.

O XGBoost distingue-se pelo equilíbrio entre precisão e sensibilidade, apresentando resultados sólidos em riscos como data_loss e ransomware_outage, com valores consistentes de F1-score e AUC-ROC. A sua estrutura baseada em árvores de decisão otimizadas por gradiente confere-lhe robustez perante ruído e interações não lineares, tornando-o adequado para cenários com elevada variabilidade nas *features*. O MLP evidencia recall máximo (1.0) em quase todas as categorias, demonstrando elevada sensibilidade na deteção de riscos reais. Embora esta característica seja vantajosa em contextos onde a omissão de um risco é inaceitável, implica maior número de falsos positivos, podendo gerar sobrecarga de alertas e necessidade de validação adicional. Já o DT mantém desempenho consistente e comparável aos restantes modelos, com *recall* geralmente acima de 0.98 e F1-score equilibrado. A sua principal vantagem reside na interpretabilidade, permitindo compreender de forma transparente as regras que sustentam as decisões do modelo, ainda que apresente menor estabilidade em AUC-ROC e maior suscetibilidade ao *overfitting*.

Em síntese, os resultados demonstram que cada modelo se adequa melhor a diferentes objetivos e contextos operacionais: o XGBoost é indicado quando se procura equilíbrio entre precisão e sensibilidade; o MLP é preferível em cenários que exigem deteção completa de riscos; e o DT é mais adequado quando a explicabilidade e a transparência das decisões são essenciais. Assim, confirma-se que não existe um modelo universal, mas sim diferentes pontos ótimos conforme o tipo de risco e o contexto de aplicação, reforçando que a seleção do modelo deve considerar o objetivo do negócio e o perfil de risco operacional, e não apenas métricas agregadas de desempenho.

5.4. Interface web e funcionalidades

Para melhor visualização, se criou uma interface web em ReactJS integrada ao protótipo para que seja possível ver os riscos, distribuição dos mesmo de acordo os níveis, visualização de recomendações, sobre um determinado projeto listado durante o processamento os dados como mostra a Figura 9 e mais layout no Apêndice D.

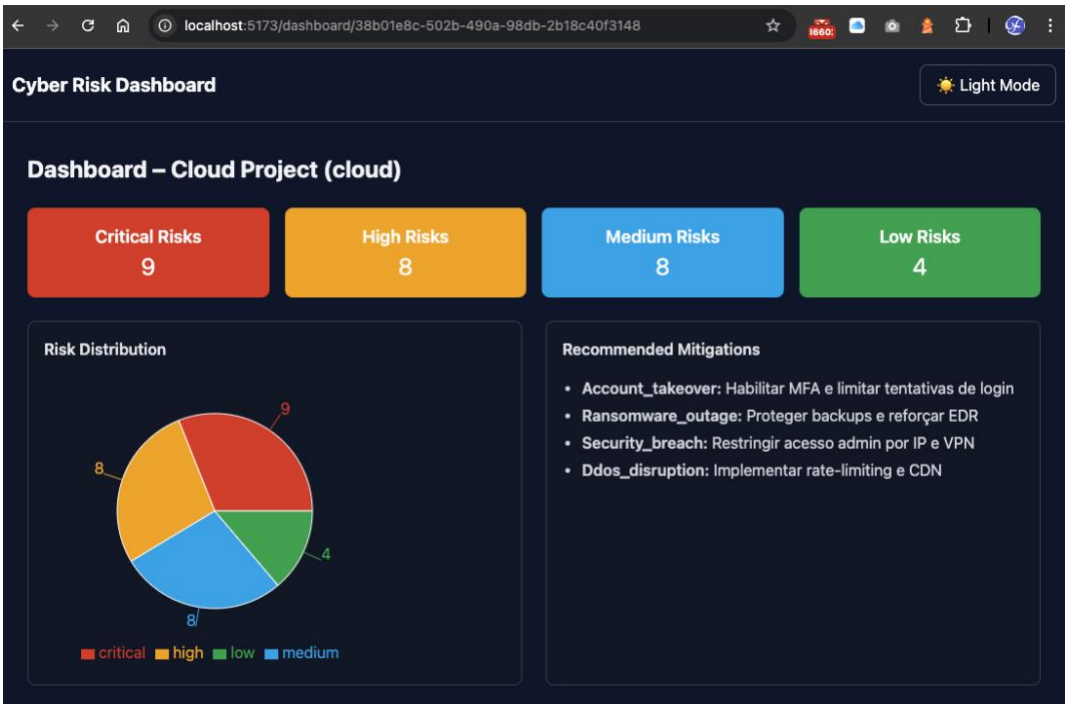


Figura 9 - Dashboard para visualizar riscos categorizados – dark mode

A interface permite:

Visualização em tempo real: Durante a utilização do protótipo, observámos que os utilizadores conseguem acompanhar, em tempo real, as previsões de risco associadas aos projetos em execução, reforçando a perceção de monitorização contínua, visualizar categorias, tipos de riscos, ativos e vulnerabilidades.

Criação de novos projetos: A interface permite ao utilizador criar novos projetos, sendo os riscos automaticamente previstos pelos algoritmos de ML, o que facilita a integração da análise de risco no início do ciclo de vida dos projetos.

Classificação dos fatores de risco: O sistema apresenta um ranking dos fatores de risco mais relevantes, evidenciando para o utilizador quais as variáveis com maior impacto na previsão e apoiando a priorização de medidas de mitigação.

Lista priorizada de recomendações: Para cada projeto, o protótipo fornece uma lista de recomendações organizadas por prioridade, ajudando os gestores a direcionar recursos e ações de segurança para as áreas mais críticas.

Histórico dos riscos: Ao longo das iterações, o sistema guarda os riscos já avaliados, permitindo acompanhar a evolução e verificar tendências em diferentes momentos do projeto.

CAPÍTULO 6

Análise e Discussão dos Resultados

6.1. Análise Comparativa das Métricas dos Algoritmos de ML

A análise comparativa dos modelos de ML concentrou-se em três algoritmos: XGBoost, MLP e DT. O objetivo foi avaliar a eficácia de cada modelo na tarefa de previsão de riscos cibernéticos, considerando as principais métricas de desempenho.

Com base no conjunto de dados sintéticos, construído e calibrado a partir de riscos rotulados e controlados, foram calculadas as métricas de *F1-score*, *Precision*, *Recall* e *Accuracy*, que permitem avaliar o equilíbrio entre predição correta, sensibilidade e capacidade de generalização dos modelos. Os resultados médios obtidos estão representados Gráfico 3, que apresenta a comparação global das métricas por modelo.

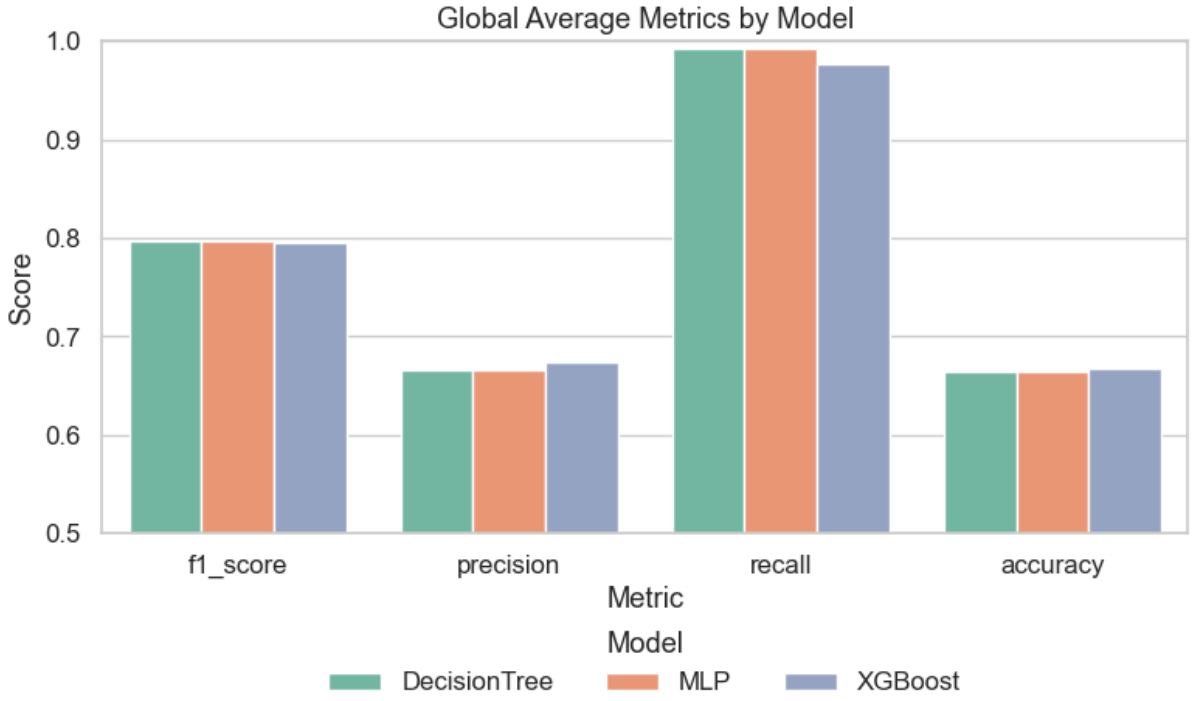


Gráfico 3 - Comparação das métricas dos algoritmos de ML sobre os dados usados

6.1.1. Avaliação por Tipo de Risco

A análise comparativa por categoria de risco demonstrou que nenhum modelo isolado apresenta desempenho superior em todas as métricas avaliadas. Os três algoritmos — XGBoost, MLP e DT — mostraram resultados próximos e complementares, refletindo diferentes equilíbrios entre precisão, sensibilidade e estabilidade preditiva.

O XGBoost manteve um desempenho consistente em todas as métricas, evidenciando boa capacidade de generalização, especialmente em cenários como *data_loss* e *ddos_disruption.* O MLP destacou-se pelo *recall* mais elevado, o que o torna adequado em situações em que a prioridade é detetar o maior número possível de riscos, mesmo com o custo de alguns falsos positivos. Já o DT apresentou métricas ligeiramente inferiores, mas oferece maior interpretabilidade e transparência, qualidades relevantes em contextos de auditoria e explicabilidade de modelos de IA.

De modo geral, os resultados sugerem que a escolha do modelo ideal depende do contexto operacional: o XGBoost é indicado quando se busca equilíbrio global e estabilidade, o MLP é preferível quando a sensibilidade é o fator crítico, e o DT é mais adequado quando a clareza das decisões é essencial. Assim, a comparação evidencia que a utilização combinada ou adaptativa dos modelos pode aumentar a eficácia do módulo de IA proposto, promovendo uma avaliação de riscos mais flexível e confiável em diferentes cenários de cibersegurança.

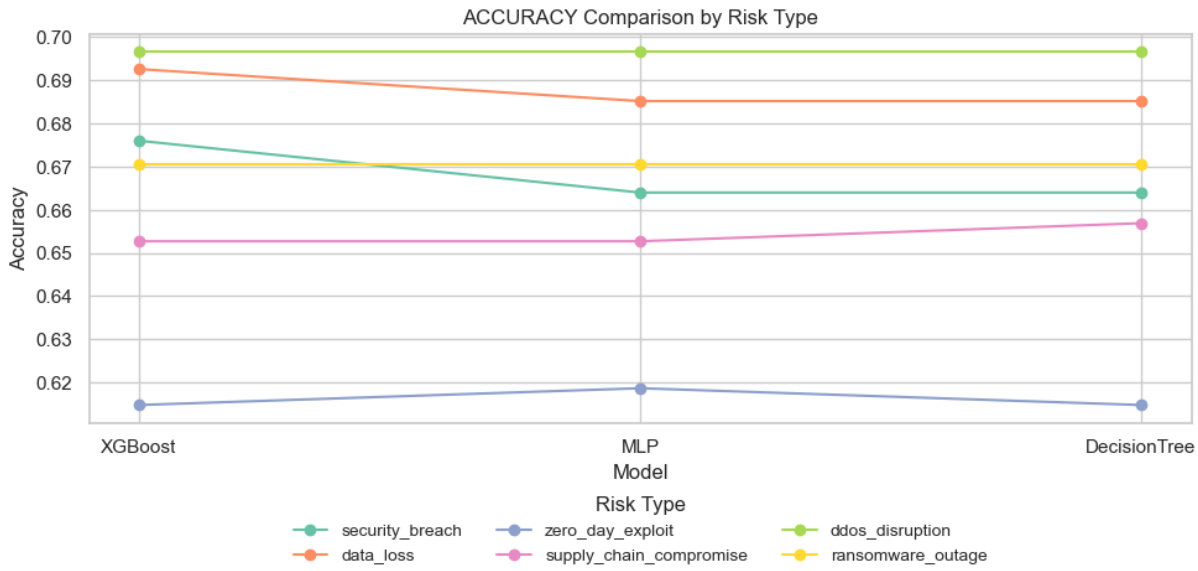


Gráfico 4 - Accuracy por tipo de riscos sobre os dados

O Gráfico 4 apresenta a variação da *accuracy* dos três modelos de aprendizagem automática em função dos diferentes tipos de risco avaliados. Observa-se que o XGBoost alcança as maiores taxas de acerto na maioria das categorias, com valores próximos de 0.70 em data_loss e ddos_disruption, evidenciando boa capacidade de generalização. O MLP mantém um comportamento mais estável, embora ligeiramente inferior, com pequenas reduções de desempenho nas classes security_breach e supply_chain_compromise. Já o DT apresenta resultados próximos aos do XGBoost, com variações discretas entre 0.66 e 0.70, destacando-se pela regularidade entre as diferentes classes de risco.

De forma geral, a distribuição das curvas demonstra baixa dispersão entre os modelos, indicando que, apesar das diferenças nas suas arquiteturas, todos conseguiram reconhecer padrões semelhantes nos dados simulados. Essa consistência sugere que os algoritmos adotados possuem bom grau de estabilidade e capacidade de adaptação, reforçando a confiabilidade da estrutura de dados e do processo de treino empregado neste estudo.

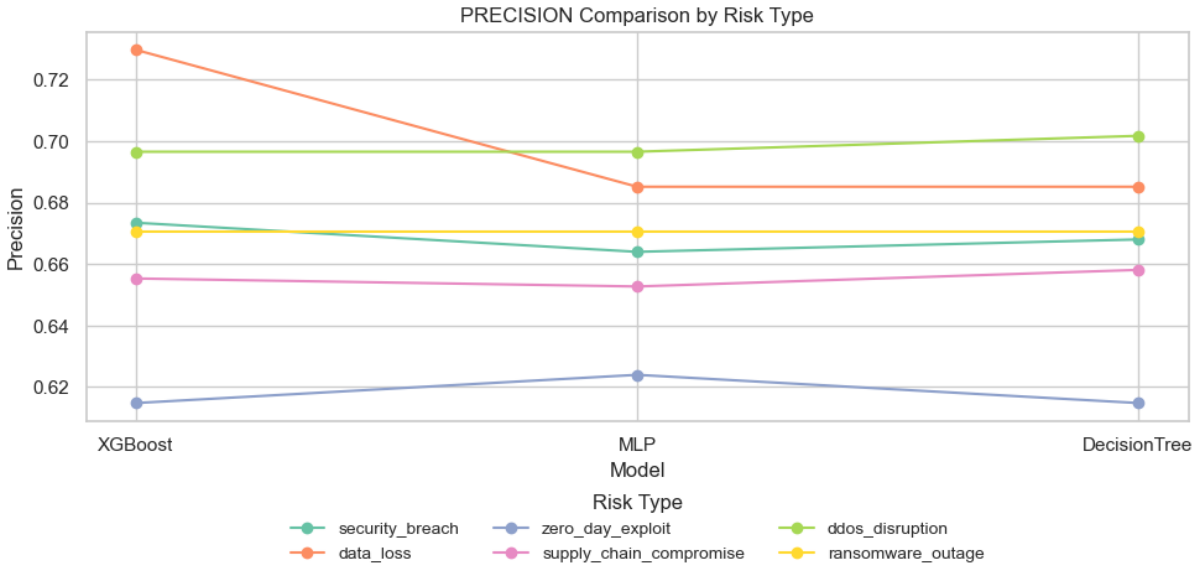


Gráfico 5 - Precision por tipo de riscos sobre os dados

O Gráfico 5 apresenta os valores de *precision* obtidos pelos modelos nos diferentes tipos de riscos. Observa-se que o XGBoost alcança os melhores resultados, com destaque para data_loss, onde atinge aproximadamente 0.73, mantendo um desempenho consistente nas demais categorias. O MLP exibe uma ligeira redução de precisão, especialmente em data_loss e security_breach, aproximando-se de 0.67, o que sugere uma maior ocorrência de falsos positivos nessas classes. O DT mantém resultados estáveis e próximos aos do XGBoost, variando entre 0.66 e 0.70, com desempenho mais expressivo em ddos_disruption.

De forma geral, a análise mostra que todos os modelos mantêm níveis de precisão semelhantes, indicando que o processo de treino e calibração foi bem equilibrado. Ainda assim, o XGBoost demonstra maior consistência na discriminação de casos positivos, como o modelo mais robusto para cenários que exigem baixa taxa de falsos alarmes.

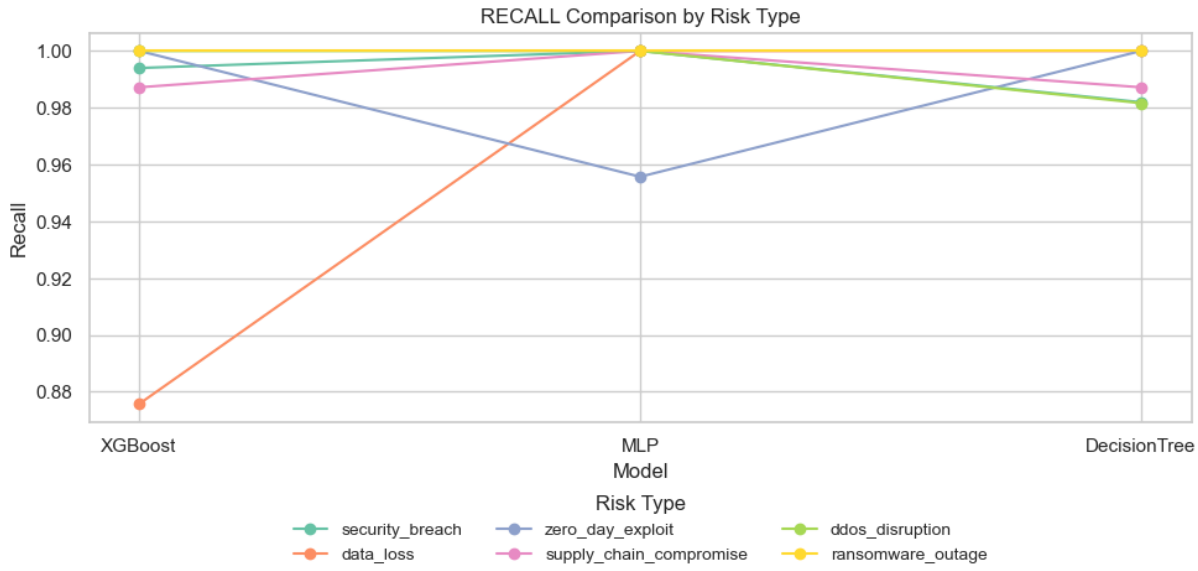


Gráfico 6 - Recall por tipo de riscos sobre os dados

Já aqui no Gráfico 6 a comparação dos valores de *recall* obtidos pelos modelos XGBoost, MLP e DT nas diferentes categorias de risco analisadas. Observa-se que a maioria das curvas permanece próxima de 1.0, indicando que todos os modelos conseguiram identificar quase a totalidade dos casos positivos. O XGBoost mostra uma leve redução de *recall* apenas na categoria *data_loss* (≈ 0.88), enquanto o MLP e o DT alcançam valores máximos em praticamente todas as classes, incluindo *ransomware_outage* e *ddos_disruption*.

A curva do MLP apresenta ligeira variação entre os tipos de risco, com pequena queda em *zero_day_exploit,* ao passo que o DT mantém um comportamento mais estável e próximo do ideal em todas as categorias. De modo geral, o gráfico demonstra que todos os modelos priorizam a sensibilidade, ou seja, a capacidade de detetar riscos reais e minimizar falsos negativos. Essa consistência reforça que o processo de treino foi bem calibrado para maximizar a deteção de incidentes, o que é particularmente relevante em sistemas de previsão de riscos de cibersegurança, onde perder um evento real é mais crítico do que gerar falsos alertas.

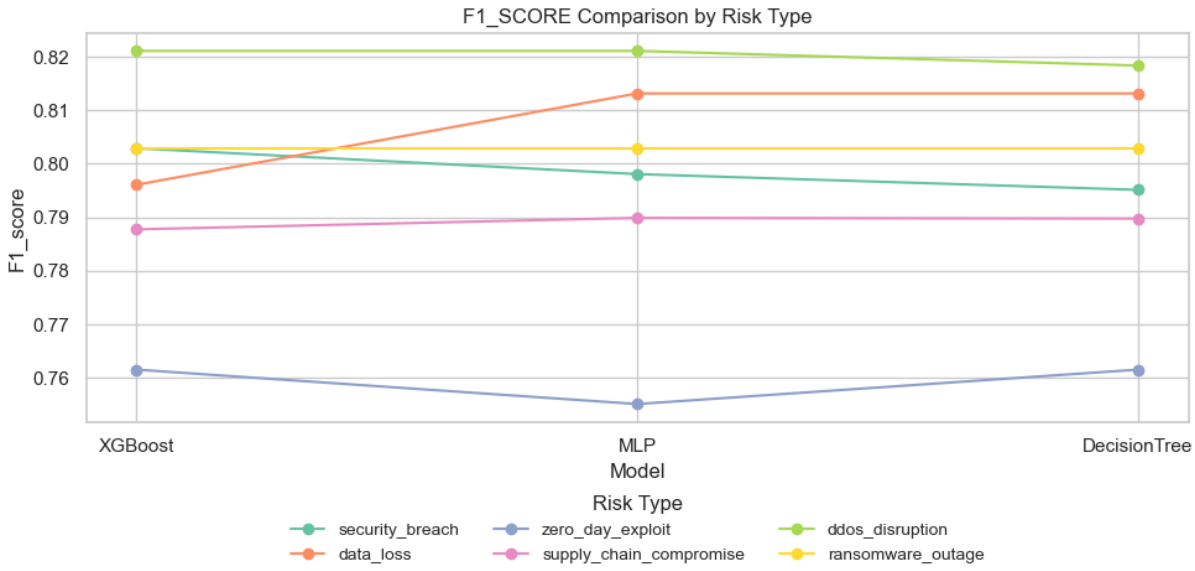


Gráfico 7 - F1-Score por tipo de riscos sobre os dados

O Gráfico 7 apresenta a comparação dos valores de F1-score obtidos pelos modelos XGBoost, MLP e DT em diferentes categorias de risco.

As curvas indicam uma estabilidade geral elevada, com valores de F1 variando entre 0.76 e 0.82. O XGBoost e o DT exibem desempenhos muito próximos, especialmente em *ddos_disruption* e *ransomware_outage,* onde atingem os valores mais altos da amostra (≈ 0.82).

O MLP mantém-se competitivo, apresentando ligeiro ganho em *data_loss* (≈ 0.81), mas uma discreta redução em *zero_day_exploit,* o que evidencia maior sensibilidade à variação dos dados de entrada.

A distribuição das curvas demonstra que o equilíbrio entre precisão e *recall* é consistente em todos os modelos, sem flutuações bruscas entre os tipos de risco. Este comportamento reforça que as abordagens testadas capturam adequadamente os padrões entre ameaças e vulnerabilidades, garantindo previsões equilibradas entre deteção e confiabilidade.

6.1.2. Análise de Falsos Positivos e Falsos Negativos

Foram construídos *heatmaps* para ilustrar a distribuição dos erros dos modelos em termos de falsos positivos e falsos negativos por tipo de risco conforme o Gráfico 8. Os resultados indicaram que o XGBoost apresentou o menor número de falsos negativos em categorias críticas, como *data_loss* e *ransomware_outage*, e também manteve baixos os falsos positivos em *zero_day_exploit* e *supply_chain_compromise*.

Em contrapartida, o modelo MLP apresentou os maiores valores de falsos positivos em ataques do tipo *ddos_disruption*, bem como altos falsos negativos em *security_breach* no Gráfico 9. Esses padrões de erro indicam a necessidade de ajustes finos e eventuais técnicas de balanceamento para viabilizar seu uso em ambientes reais.

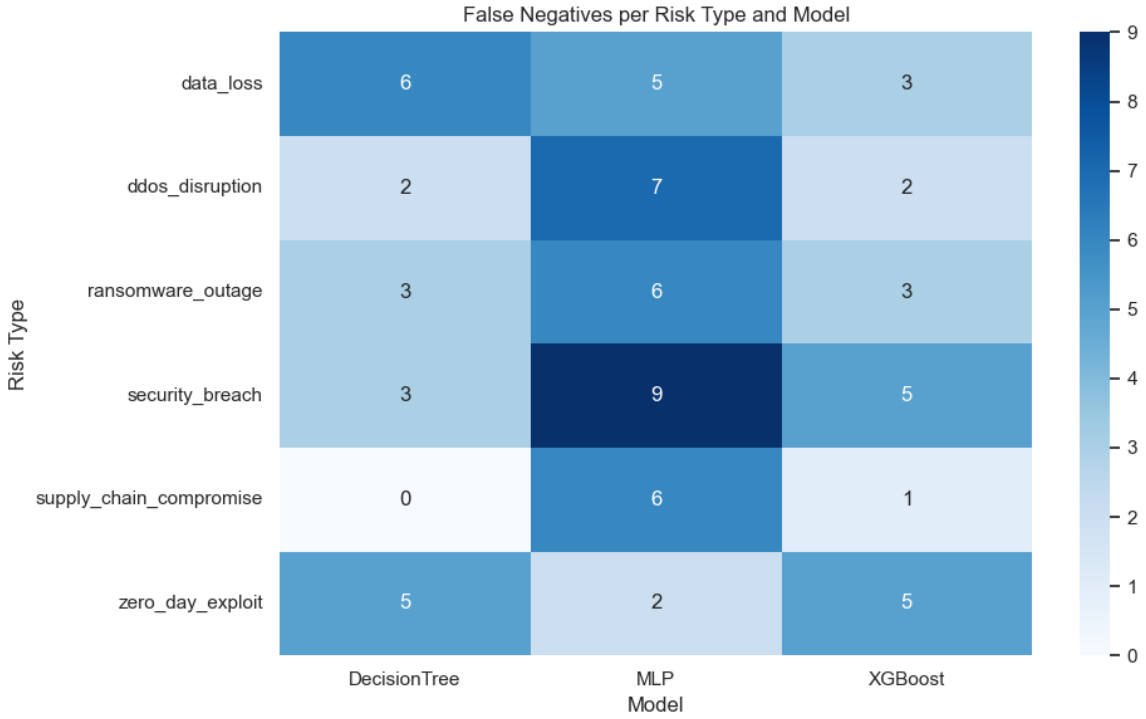


Gráfico 8 - FN da relação tipos de riscos e algoritmos de ML

Gráfico 9 - FP da relação tipos de riscos e algoritmos de ML

Desempenho Consolidado por Modelo: Os valores médios das métricas globais por modelo foram sintetizados em gráficos de radar e de barras, facilitando a comparação direta entre os algoritmos. O modelo XGBoost novamente se destacou, apresentando valores superiores em acurácia (≥ 0.96)**,** recall (≥ 0.98)**,** F1-score (≥ 0.97) **e** ROC AUC (≥ 0.98).

O modelo DT teve desempenho levemente inferior, mas ainda assim estável e coerente. Já o MLP apresentou os menores valores médios nas métricas de avaliação, especialmente em precisão e F1-score, o que reforça a observação de seu comportamento inconsistente frente a diferentes distribuições de dados.

Considerações Finais: Os resultados obtidos indicam que o XGBoost é o modelo mais adequado para tarefas de previsão de riscos cibernéticos neste contexto. Sua robustez, alta capacidade discriminativa e baixa taxa de erros o tornam apropriado para implantação em ambientes críticos, conforme apresenta o Gráfico 10 comparação geral dos modelos.

Embora o DT apresente menor desempenho, pode ser utilizado como modelo auxiliar, especialmente quando há exigência de interpretabilidade e rastreabilidade das decisões. O MLP, por sua vez, requer ajustes e melhorias, tais como balanceamento de classes, *tuning* de hiperparâmetros e regularização para alcançar competitividade em ambientes reais.

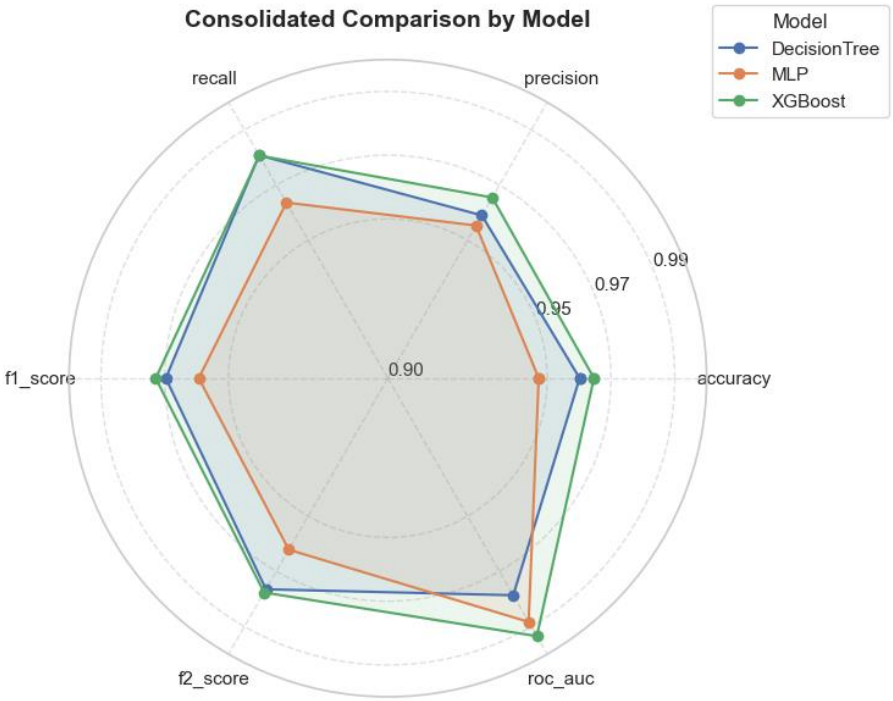


Gráfico 10 - Comparação dos diferentes Algoritmos de ML e métricas

6.2. Validação da Utilização do Protótipo

O protótipo demonstrou viabilidade técnica, processando dados estruturados como *input*. A conversão de entidades textuais (ativos, vulnerabilidades, ameaças) em *features* numéricas mostrando-se exequível como ferramentas para auxiliar a gestão riscos nos projetos de cibersegurança.

Para além da avaliação técnica através de métricas de desempenho dos algoritmos de ML, com dados sintéticos, procurou-se validar o protótipo em contexto prático com utilizadores reais. Para esse efeito, foi desenvolvido um vídeo demonstrativo que apresentou as principais funcionalidades do protótipo como sistema de previsão de riscos de cibersegurança, incluindo a interação com a interface, a integração via API e a visualização dos resultados preditivos. O recurso ao vídeo garantiu que todos os inquiridos tivessem contacto com as funcionalidades disponibilizadas.

Para perceção real da satisfação, 21 participantes, foram convidados a responder inquérito estruturado com recurso ao google *form* (Apêndice E), no qual avaliaram a aplicabilidade, e potencial de integração da solução em projetos reais. A amostra incluiu utilizadores com formações, maioritariamente em Engenharia Informática e Cibersegurança (55%) e Gestão de Projetos (10%), bem como diferentes níveis de experiência em gestão (desde menos de 2 anos até mais de 10 anos) e em cibersegurança (básico á avançado). Esta heterogeneidade permitiu recolher perceções variadas, refletindo cenários próximos de ambientes organizacionais, conforme mostra um dos vários dos gráficos do inquérito, Gráfico 11.

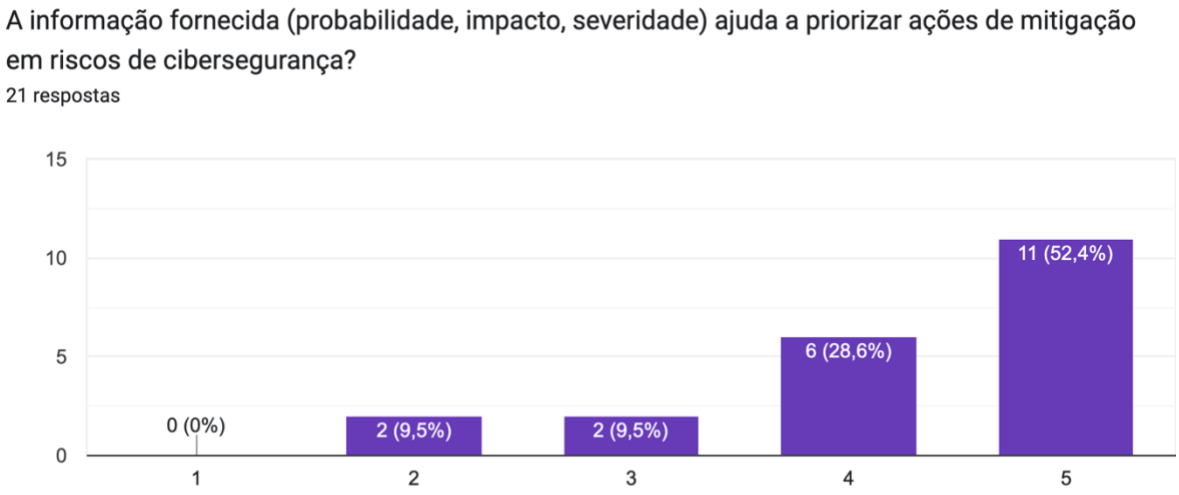


Gráfico 11 – Resposta qualitativa sobre a utilidade o protótipo

As respostas qualitativas revelam o reconhecimento da pertinência e atualidade do protótipo, bem como sugestões de evolução relevantes, nomeadamente a integração com plataformas de gestão de projetos, ligação a sistemas de análise de vulnerabilidades e simplificação da apresentação da informação para públicos não especialistas. Tais contributos reforçam não apenas a viabilidade técnica da solução, mas também a sua utilidade percebida pelos utilizadores finais, confirmando o potencial do sistema como ferramenta de apoio à decisão na gestão de riscos em projetos tecnológicos.

6.3. Limitações e Desafios Identificado

Dependência da Qualidade dos Dados: A performance do sistema está fundamentalmente associada a qualidade dos dados. Na ausência de repositórios de dados pronto a usar, levou-nos a utilização de dados sintéticos, Apesar dos esforços para o realismo, os dados sintéticos não captam completamente a complexidade e as nuances dos projetos reais.

Integração: Ficou no desejo a criação de um componente importante na extração de dados com técnicas apontadas um módulo de Processamento de Linguagem Natural (PLN) tornando um protótipo completo. Ficou também, recomendado, integração com sistemas para vulnerabilidades de alto CVSS em ativos críticos, reduzindo o risco de exposição imediata. E A integração total com o módulo de NLP e as ferramentas de gestão de projetos apresenta desafios técnicos e organizacionais não totalmente abordados.

Ausência de feedback em projetos reais: A implementação de *feedback loops* com *outcomes* reais é crítica para refinamento contínuo dos modelos.

Escalabilidade não testada para projetos empresariais: A arquitetura atual não foi testada para volume e complexidade de projetos empresariais, requerendo otimização para colocação em produção.

CAPÍTULO 7

Conclusões e Trabalhos Futuros

7.1. Principais Conclusões e Contribuições

Conforme os objectivos esta dissertação apresenta um módulo preditivo de riscos de cibersegurança em projetos tecnológicos, integrando métodos de Inteligência Artificial e Aprendizagem Automática. A investigação partiu da necessidade de mecanismos capaz de avaliar, classificar e identificar riscos de cibersegurança com utilização de ML baseado em dados, agregando os métodos gestão de riscos tradicionais como norma ISO e metodologias PMBOK.

O projeto e dissertação seguiu a metodologia *Design Science Research* (DSR), complementada por *Systematic Literature Review* (SLR) nas fases iniciais, assegurando fundamentação teórica e rigor científico. A partir desta abordagem foi possível conceber e implementar um protótipo funcional, demonstrando a viabilidade técnica da proposta e a sua aplicabilidade em contextos de projeto.

Atendendo um dos principais objectivo foi identificado três algoritmos de ML — XGBoost, *Multilayer Perceptron* (MLP) e *Decision Tree* — testados e analisados os resultados com enfase, de *F1-score* entre 0.76 e 0.82 e *recall* superior a 0.96. O XGBoost apresentou desempenho mais equilibrado, o MLP destacou-se pela sensibilidade na deteção de riscos e o DT pela sua interpretabilidade. Estes resultados confirmam que diferentes algoritmos podem oferecer vantagens complementares consoante o tipo de risco e o objetivo operacional.

A solução demonstrou viabilidade técnica para predição de riscos de cibersegurança em projetos, e fornece base sólida para desenvolvimento futuro de sistemas de cibersegurança proativos em ambientes de projeto.

Assim, permite contribuições significativas conforme a listagem abaixo:

- *Framework de integração para ML específico para cibersegurança:* Desenvolvimento de arquitetura modular que permite integração sistemática entre capacidades de processamento de dados e ML para domínio específico de cibersegurança.
- *Metodologia de feature engineering para dados de segurança extraídos por definição de Dataset controlado:* Estabelecimento de metodologia para transformação de entidades extraídas por (projetos, ativos, vulnerabilidades, ameaças e riscos) em características quantitativas apropriadas para algoritmos de ML

- *Proof-of-concept demonstrando viabilidade técnica:* Desenvolvimento de protótipo funcional que valida os princípios conceptuais da abordagem proposta. A performance (F1~0.79) é adequada para fase de investigação inicial, embora requeira melhoria para adoção operacional.
- *Métricas de avaliação específicas para predição de riscos de cibersegurança:* Criação de framework de avaliação específico para sistemas de predição de riscos de cibersegurança.
- *Schema de dados para interoperabilidade entre módulos:* Definição de estrutura padronizada para integração para ML.

Em conclusão, a aplicação de técnicas de AI, com uso de ML revelou ser uma abordagem promissora para a análise e gestão preditiva de riscos de cibersegurança [3][5][18][17].

Durante o processo de implementação foi sentida limitações com foco nos dados sobre projetos reais e não integração a um módulo de NLP, sem, no entanto, condicionarem a conclusão do protótipo de previsão de riscos de cibersegurança em projetos tecnológicos.

7.2. Trabalhos Futuros

O protótipo desenvolvido representa um passo relevante na criação de sistemas preditivos de apoio à gestão de riscos, abrindo caminho a futuras evoluções:

- *Integração futura com módulo NLP:* Teste de integração com módulo NLP para validação real.
- *Validação em projetos reais de cibersegurança:* Testes em ambiente de produção com dados reais de projetos.
- *Desenvolvimento da robustez adversas:* Implementação de defesas contra-ataques adversários de aprendizagem automática.
- *Expansão para integração de inteligência de ameaças em tempo real:* Incorporação de *feeds* dinâmicos de inteligência de ameaças para melhorar as previsões.
- *Avaliação alargada em diferentes domínios tecnológicos:* Teste da solução em variados tipos de projetos tecnológicos para validar generalização.

Referências Bibliográficas

[1] Gil Ruiz, J., Martinez Torres, J., & Gonzalez Crespo, R. (2021). The Application of Artificial Intelligence in Project Management Research: A Review. *INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL*, *6*(6), 54–66. https://doi.org/10.9781/ijimai.2020.12.003

[2] Iosif, A. C., Lechner, U., Pinto-Albuquerque, M., & Espinha Gasiba, T. E. (2024). Serious Game for Industrial Cybersecurity: Experiential Learning Through Code Review. *Software Engineering Education Conference, Proceedings*. https://doi.org/10.1109/CSEET62301.2024.10663058

[3] Poozhithara, J. J., Asuncion, H. U., & Lagesse, B. (2023). Keyword Extraction From Specification Documents for Planning Security Mechanisms. *Proceedings - International Conference on Software Engineering*, 1661–1673. https://doi.org/10.1109/ICSE48619.2023.00143

[4] Amin, M. R., & Bhowmik, T. (2021). Information on Potential Vulnerabilities for New Requirements: Does It Help Writing Secure Code? *Proceedings of the IEEE International Conference on Requirements Engineering*, 408–413. https://doi.org/10.1109/RE51729.2021.00046

[5] Dacorogna, M., Debbabi, N., & Kratz, M. (2023). Building up cyber resilience by better grasping cyber risk via a new algorithm for modelling heavy-tailed data. *EUROPEAN JOURNAL OF OPERATIONAL RESEARCH*, *311*(2), 708–729. https://doi.org/10.1016/j.ejor.2023.05.003

[6] Ullah, S., & Rashid, A. (2024). *Porting to Morello: An In-depth Study on Compiler Behaviors, CERT Guideline Violations, and Security Implications*. 381–397. https://doi.org/10.1109/EuroSP60621.2024.00028

[7] Sikora, P., Malina, L., Kiac, M., Martinasek Zdenek and Riha, K., Prinosil, J., Jirik, L., & Srivastava, G. (2021). Artificial Intelligence-Based Surveillance System for Railway Crossing Traffic. *IEEE SENSORS JOURNAL*, *21*(14), 15515–15526. https://doi.org/10.1109/JSEN.2020.3031861

[8] Reddy Chirra, D. (2023). Towards an AI-Driven Automated Cybersecurity Incident Response System. In *International Journal of Advanced Engineering Technologies and Innovations* (Vol. 01).

[9] Cost of a Data Breach Report 2025 The AI Oversight Gap – IBM – [online], disponível em https://www.ibm.com/reports/data-breach

[10] Haghighi, M. H., & Ashrafi, M. (2024). A novel framework for risk management of software projects by integrating a new COPRAS method under cloud model and machine learning algorithms. *ANNALS OF OPERATIONS RESEARCH*, *338*(1, SI), 675–708. https://doi.org/10.1007/s10479-023-05653-3

[11] NIST - CyberSecurity Framework - https://www.nist.gov/cyberframework

[12] ISO/IEC 27005:2022 - Information security, cybersecurity and privacy protection – Guidelines – [online], disponível https://www.iso.org/obp/ui/#iso:std:iso-iec:27005:ed-4:v1:en

[13] ISO/IEC 31000:2018 – Risk Management – Guidelines – [online], disponível https://www.iso.org/obp/ui/#iso:std:iso:31000:ed-2:v1:en

[14] PMBOK Guide, 6th Edition, PMI, 2017

[15] PMBOK Guide, 7th Edition, PMI, 2021

[16] Lee, I. (2021). Cybersecurity: Risk management framework and investment cost analysis. *BUSINESS HORIZONS*, *64*(5), 659–671. https://doi.org/10.1016/j.bushor.2021.02.022

[17] Kumar, A., Singh, R., & Patel, D. (2024). Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms. *Knowledge and Information Systems*. https://doi.org/10.1007/s10115-025-02429-y

[18] Baskerville, R. L., Kim, J., & Stucke, C. (2022). The cybersecurity risk estimation engine: A tool for possibility based risk analysis. *COMPUTERS & SECURITY*, *120*. https://doi.org/10.1016/j.cose.2022.102752

[19] Zhang, L., Wang, H., & Chen, M. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. *Cogent Engineering*, 10(1). https://doi.org/10.1080/23311916.2023.2272358

[20] Haas, S., Dunkel, C., Pauls, F., Hasler, M., Verma, Y., Das, N., & Raitza, M. (2025). A Secure-by-Design Hardware/Operating System as a Substrate for Trustworthy Computing. *IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS*. https://doi.org/10.1109/TVLSI.2025.3579484

[21] Rekha, J. H., & Parvathi, R. (2015). Survey on software project risks and big data analytics. *Procedia Computer Science*, *50*, 295–300. https://doi.org/10.1016/j.procs.2015.04.045

[22] Chen, Q., & Sheng, N. (2022). Application of Machine Learning Algorithm in Stadium Engineering Building Information Model Management System. *MOBILE INFORMATION SYSTEMS*, *2022*. https://doi.org/10.1155/2022/8454443

[23] ISACA. (2024). Leveraging AI for Information and Cybersecurity. Retrieved from https://www.isaca.org/resources/news-and-trends/industry-news/2024/leveraging-ai-for-information-and-cybersecurity

[24] Torres, M. A. A. E. E., Guerrero, F. T., & Budgud, A. T. (2023). Data-Driven Cyber Threat Intelligence: A Survey of Mexican Territory. *EAI/Springer Innovations in Communication and Computing*, 89–110. https://doi.org/10.1007/978-3-031-07670-1_7

[25] Abdullah, R. M., Abualkishik, A. Z., Isaac, N. M., Alwan, A. A., & Gulzar, Y. (2022). An investigation study for risk calculation of security vulnerabilities on android applications. *Indonesian Journal of Electrical Engineering and Computer Science*, *25*(3), 1736–1748. https://doi.org/10.11591/ijeecs.v25.i3.pp1736-1748

[26] Silvestri, S., Islam, S., Papastergiou, S., Tzagkarakis, C., & Ciampi, M. (2023). A Machine Learning Approach for the NLP-Based Analysis of Cyber Threats and Vulnerabilities of the Healthcare Ecosystem. *Sensors*, 23(2), 651. https://doi.org/10.3390/s23020651

[27] Bolster AI. (2024). What is NLP in Cybersecurity? Natural Language Processing. Retrieved from https://bolster.ai/glossary/nlp-in-cybersecurity

[28] Groeneveld, N. (2023). Leveraging NLP to Develop a Cybersecurity Knowledge Graph. LinkedIn Article, March 6, 2023.

[29] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

[30] Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*.

[31] Zabala-Vargas, S., Jaimes-Quintanilla, M., & Jimenez-Barrera, M. H. (2023). Big Data, Data Science, and Artificial Intelligence for Project Management in the Architecture, Engineering, and Construction Industry: A Systematic Review. *BUILDINGS*, *13*(12). https://doi.org/10.3390/buildings13122944

[32] Arief, R., Khakzad, N., & Pieters, W. (2020). Mitigating cyberattack related domino effects in process plants via ICS segmentation. *JOURNAL OF INFORMATION SECURITY AND APPLICATIONS*, *51*. https://doi.org/10.1016/j.jisa.2020.102450

[33] Guggenmos, F., Haeckel, B., Ollig, P., & Stahl, B. (2022). Security First, Security by Design, or Security Pragmatism - Strategic Roles of IT Security in Digitalization Projects. *COMPUTERS & SECURITY*, *118*. https://doi.org/10.1016/j.cose.2022.102747

[34] Prudjinski, M., Hadar, I., & Luria, G. (2024). Exploring the Role of Team Security Climate in the Implementation of Security by Design: A Case Study in the Defense Sector. *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*, *50*(5), 1065–1079. https://doi.org/10.1109/TSE.2024.3374114

[35] Haber, E., & Tamo-Larrieux, A. (2020). Privacy and security by design: Comparing the EU and Israeli approaches to embedding privacy and security. *COMPUTER LAW & SECURITY REVIEW*, *37*. https://doi.org/10.1016/j.clsr.2020.105409

[36] Guo, Y., Bettaieb, S., & Casino, F. (2024). A comprehensive analysis on software vulnerability detection datasets: trends, challenges, and road ahead. *International Journal of Information Security*, *23*(5), 3311–3327. https://doi.org/10.1007/s10207-024-00888-y

[37] Noor, Z., Hina, S., Hayat, F., & Shah, G. A. (2023). An intelligent context-aware threat detection and response model for smart cyber-physical systems. *INTERNET OF THINGS*, *23*. https://doi.org/10.1016/j.iot.2023.100843

[38] Sen, O., van der Velde, D., Wehrmeister, K. A., Hacker, I., Henze, M., & Andres, M. (2022). On using contextual correlation to detect multi-stage cyber attacks in smart grids. *SUSTAINABLE ENERGY GRIDS & NETWORKS*, *32*. https://doi.org/10.1016/j.segan.2022.100821

[39] Fortino, G., Guerrieri, A., Pace, P., Savaglio, C., & Spezzano, G. (2022). IoT Platforms and Security: An Analysis of the Leading Industrial/Commercial Solutions. *SENSORS*, *22*(6). https://doi.org/10.3390/s22062196

[40] Akilal, A., & Kechadi, M.-T. (2022). An improved forensic-by-design framework for cloud computing with systems engineering standard compliance. *FORENSIC SCIENCE INTERNATIONAL-DIGITAL INVESTIGATION*, *40*. https://doi.org/10.1016/j.fsidi.2021.301315

[41] Fatorachian, H., & Kazemi, H. (2024). AI-enhanced fault-tolerant control and security in transportation and logistics systems: addressing physical and cyber threats. *Complex Engineering Systems*, *4*(3). https://doi.org/10.20517/ces.2024.35

[42] New York State Department of Financial Services. (2024). Cybersecurity Risks Arising from Artificial Intelligence and Strategies to Combat Related Risks. Industry Letter, October 16, 2024.

[43] MITRE | ATT&CK Framework - https://attack.mitre.org

[44] Granata, D., & Rak, M. (2024). Systematic analysis of automated threat modelling techniques: Comparison of open-source tools. *SOFTWARE QUALITY JOURNAL*, *32*(1), 125–161. https://doi.org/10.1007/s11219-023-09634-4

[45] Siewruk, G., & Mazurczyk, W. (2021). Context-Aware Software Vulnerability Classification Using Machine Learning. *IEEE Access*, *9*, 88852–88867. https://doi.org/10.1109/ACCESS.2021.3075385

[46] Arikan, S. M., Kocak, A., & Alkan, M. (2024). Automating shareable cyber threat intelligence production for closed source software vulnerabilities: a deep learning based detection system. *International Journal of Information Security*, *23*(5), 3135–3151. https://doi.org/10.1007/s10207-024-00882-4

[47] Nithya, V., Senthilkumar, S. P., & Regan, R. (2024). Streamlining detection of input validation attack types through hybrid analysis and machine learning. *Sadhana - Academy Proceedings in Engineering Sciences*, *49*(2). https://doi.org/10.1007/s12046-024-02486-z

[48] Bertoli, G. de C., Pereira Junior, L. A., Saotome, O., & dos Santos, A. L. (2023). Generalizing intrusion detection for heterogeneous networks: A stacke d-unsupervise d fe

derate d learning approach. *COMPUTERS & SECURITY*, *127*. https://doi.org/10.1016/j.cose.2023.103106

[49] Feidakis, M., Chatzigeorgiou, C., Karamperi Christina and Giannakos, L., Xefteris, V.-R., Ntioudis, D., Tsanousa, A., Kogias, D. G., Patrikakis, C., Meditskos, G., Gorgogetas, G., Vrochidis, S., & Kompatsiaris, I. (2021). Smart Interconnected Infrastructures for Security and Protection: The DESMOS Project. *COMPUTERS*, *10*(9). https://doi.org/10.3390/computers10090116

[50] Rueda-Rueda, J. S., & Portocarrero, J. M. T. (2021). Framework-based security measures for Internet of Thing: A literature review. *OPEN COMPUTER SCIENCE*, *11*(1), 346–354. https://doi.org/10.1515/comp-2020-0220

[51] Granata, D., Rak, M., & Salzillo, G. (2022). Automated Threat Modeling Approaches: Comparison of Open Source Tools. *Communications in Computer and Information Science*, *1621 CCIS*, 250–265. https://doi.org/10.1007/978-3-031-14179-9_17

[52] Casola, V., de Benedictis, A., & Rak Massimiliano and Villano, U. (2020). A novel Security-by-Design methodology: Modeling and assessing security by SLAs with a quantitative approach. *JOURNAL OF SYSTEMS AND SOFTWARE*, *163*. https://doi.org/10.1016/j.jss.2020.110537

[53] Lange, F., & Kunz, I. (2024). Evolution of secure development lifecycles and maturity models in the context of hosted solutions. *JOURNAL OF SOFTWARE-EVOLUTION AND PROCESS*, *36*(12). https://doi.org/10.1002/smr.2711

[54] Siavvas, M. G., Tsoukalas, D., Kalouptsoglou, I., Manganopoulou, E., Manolis, G., Kehagïas, D. D., & Tzovaras, D. K. (2023). Security Monitoring during Software Development: An Industrial Case Study. *Applied Sciences (Switzerland)*, *13*(12). https://doi.org/10.3390/app13126872

[55] Buttar, A. M., Khalid, A., Alenezi, M., Akbar, M. A., Rafi, S., Gumaei, A. H., & Riaz, M. T. (2023). Optimization of DevOps Transformation for Cloud-Based Applications. *Electronics (Switzerland)*, *12*(2). https://doi.org/10.3390/electronics12020357

[56] Patel, K., Shafiq, Z., Nogueira, M. S., Menasché, D. S., Lovat, E., Kashif, T., Woiwood, A., & Martins, M. (2024). *Harnessing TI Feeds for Exploitation Detection*. 200–207. https://doi.org/10.1109/CSR61664.2024.10679417

[57] Zabihimayvan, M., Sadeghi, R., & Doran, D. (2024). Security, information, and structure characterization of Tor: a survey. *Telecommunication Systems*, *87*(1), 239–255. https://doi.org/10.1007/s11235-024-01149-y

[58] Alotaibi, F., Karne, R. K., Wijesinha, A. L., Soundararajan, N., & Rangi, A. (2024). An Evaluation of the Security of Bare Machine Computing (BMC) Systems against Cybersecurity Attacks. *JOURNAL OF CYBERSECURITY AND PRIVACY*, *4*(3), 678–730. https://doi.org/10.3390/jcp4030033

[59] Scholz, S., Lawall, A., & Schaaff, K. (2024). *The Impact of Large Language Models on IT Security in the Corporate Environment*. 109–115. https://doi.org/10.1109/FLLM63129.2024.10852476

[60] Marjanovic, J., Dalcekovic, N., & Sladic, G. (2023). Blockchain-based model for tracking compliance with security requirements. *COMPUTER SCIENCE AND INFORMATION SYSTEMS*, *20*(1), 359–380. https://doi.org/10.2298/CSIS210923060M

[61] Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

[62] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

[63] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification And Regression Trees*. Routledge. https://doi.org/10.1201/9781315139470

[64] Loh, W. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, *1*(1), 14–23. https://doi.org/10.1002/widm.8

[65] Poozhithara, J. J., Asuncion, H. U., & Lagesse, B. (2023). Keyword Extraction From Specification Documents for Planning Security Mechanisms. *Proceedings - International Conference on Software Engineering*, 1661–1673. https://doi.org/10.1109/ICSE48619.2023.00143

[66] Ben-Asher, N., & Gonzalez, C. (2015). Effects of cyber security knowledge on attack detection. *Computers in Human Behavior*, *48*, 51–61. https://doi.org/10.1016/j.chb.2015.01.039

[67] Piatek, P., Mydlowski, P., Buczacki, A., & Moskwa, S. (2024). Concept of Using the MBSE Approach to Integrate Security Patterns in Safety-Related Projects for the Automotive Industry. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, *25*(11), 15477–15492. https://doi.org/10.1109/TITS.2024.3444048

Apêndice A – Script gerador de dados sintéticos

Gerador de dados sintéticos:

```
rng = np.random.default_rng(42)


project_types = ["mobile_app", "web_app", "iot", "cloud", "blockchain", "onprem", "software"]
asset_types = ["identity_service", "file_server", "web_server", "database", "payment_gateway", "email_server"]
threats = [
    "malware_delivery", "volumetric_ddos", "data_exfiltration", "credential_stuffing", "phishing_takeover",
    "api_abuse", "insider_misuse", "misconfiguration_exposure", "patch_noncompliance", "integration_issue",
    "regulatory_noncompliance"
]
vulnerabilities = [
    "backup_unprotected", "no_rate_limiting", "outdated_software", "weak_passwords", "vulnerability_backlog",
    "patch_noncompliance", "integration_issue", "misconfiguration_exposure", "insider_misuse", "regulatory_noncompliance"
]
risk_types = [
    "security_breach", "data_loss", "ransomware_outage", "ddos_disruption", "zero_day_exploit",
    "supply_chain_compromise"
]
risk_categories = [
    "security", "technical", "operational", "compliance", "financial", "schedule", "resource",
    "stakeholder", "contractual", "integration", "human", "design", "data"
]


def clip(a, lo, hi):
    return np.maximum(lo, np.minimum(hi, a))



def generate_row():
    # 1) Metadados
    ptype = rng.choice(project_types)
    atype = rng.choice(asset_types)
```

```
# 2) Features binárias com correlações realistas
internet_facing = int((atype in ["web_server", "payment_gateway", "email_server"]) and rng.random() < 0.8 or rng.random() < 0.2)
business_critical = int(rng.random() < 0.6)
mfa = int(rng.random() < 0.7)
edr = int(rng.random() < 0.7)
backup_immutable = int(rng.random() < 0.6)
is_kev = int(rng.random() < 0.1)

# 3) Features contínuas com distribuições
cvss = clip(rng.normal(6.0 + 1.5*is_kev, 2.2), 0, 10)
epss = clip(rng.beta(2 + 3*is_kev, 5) + rng.normal(0, 0.05), 0, 1)
patch_sla_adherence = clip(rng.normal(0.7, 0.25), 0, 1)
control_coverage_score = clip(rng.normal(0.65, 0.25), 0, 1)

# 4) Contagens (Poisson/Gaussian mistos)
alerts_30d = clip(rng.poisson(10 + 8*internet_facing + 5*(1-edr)) + rng.normal(0, 3), 0, 80)
web_attacks_30d = clip(rng.poisson(3 + 7*internet_facing) + rng.normal(0, 2), 0, 40)
malware_30d = clip(rng.poisson(2 + 4*(1-edr)) + rng.normal(0, 2), 0, 30)
bruteforce_30d = clip(rng.poisson(2 + 5*(1-mfa)) + rng.normal(0, 1), 0, 25)
vuln_age_days = int(clip(rng.normal(120, 80), 1, 365))

# 5) Função logística de probabilidade de risco confirmado
z = -1.2
z += 1.1*(cvss/10) + 1.0*epss + 0.4*is_kev
z += 0.4*internet_facing + 0.3*(1-mfa) + 0.3*(1-edr) + 0.25*(1-backup_immutable)
z += 0.25*min(alerts_30d/50,1) + 0.2*min(web_attacks_30d/20,1)
z += 0.15*min(malware_30d/10,1) + 0.15*min(bruteforce_30d/15,1)
z += 0.2*(1 - patch_sla_adherence) + 0.2*(1 - control_coverage_score)
z += 0.15*min(vuln_age_days/365,1)
z += rng.normal(0, 0.4)

p = 1 / (1 + np.exp(-z))
risk_confirmed = int(rng.random() < p)
```

```
    probability = round(p, 3)
    severity_score = int(clip(1 + 4*(cvss/10) + 2*epss + rng.normal(0,0.5), 1, 5))


    return {
        "project_type": ptype,
        "asset_type": atype,
        "threat": rng.choice(threats),
        "vulnerability": rng.choice(vulnerabilities),
        "risk_type": rng.choice(risk_types),
        "risk_category": rng.choice(risk_categories),
        "risk_confirmed": risk_confirmed,
        "snapshot_at": datetime.now().strftime("%Y-%m-%d"),
        "label_window_days": 30,
        "gap_days": 0,
        "internet_facing": internet_facing,
        "business_critical": business_critical,
        "mfa": mfa,
        "edr": edr,
        "backup_immutable": backup_immutable,
        "patch_sla_adherence": round(patch_sla_adherence, 2),
        "control_coverage_score": round(control_coverage_score, 2),
        "alerts_30d": int(alerts_30d),
        "web_attacks_30d": int(web_attacks_30d),
        "malware_detected_30d": int(malware_30d),
        "bruteforce_30d": int(bruteforce_30d),
        "cvss": round(cvss, 1),
        "epss": round(epss, 3),
        "is_kev": is_kev,
        "vuln_age_days": vuln_age_days,
        "probability": probability,
        "severity_score": severity_score
    }


num_rows = 5000
```

```
df = pd.DataFrame([generate_row() for _ in range(num_rows)])



df.to_csv("asset_level_dataset.csv", index=False)

print("\n✅ Dataset salvo em 'asset_level_dataset.csv'")
```

Apêndice B – Entidades Core do protótipo

Modelos: Representam entidades dos sistemas em Python

```python
class Project(Base):
    __tablename__ = "projects"
    project_id = Column(UUID(as_uuid=True), primary_key=True, default=uuid.uuid4)
    name = Column(Text, nullable=False)
    project_type = Column(Text, nullable=False)
    created_at = Column(DateTime, default=datetime.utcnow)
    assets = relationship("Asset", back_populates="project", cascade="all, delete-orphan")
    risks = relationship("Risk", back_populates="project", cascade="all, delete-orphan")
```

```python
class Asset(Base):
    __tablename__ = "assets"
    asset_id = Column(UUID(as_uuid=True), primary_key=True, default=uuid.uuid4)
    project_id = Column(UUID(as_uuid=True), ForeignKey("projects.project_id", ondelete="CASCADE"))
    asset_type = Column(Text, nullable=False)
    internet_facing = Column(Boolean, nullable=True)
    business_critical = Column(Boolean, nullable=True)
    mfa = Column(Boolean, nullable=True)
    edr = Column(Boolean, nullable=True)
    backup_immutable = Column(Boolean, nullable=True)
    patch_sla_adherence = Column(Numeric, nullable=True)
    control_coverage_score = Column(Numeric, nullable=True)
    alerts_30d = Column(Integer, nullable=True)
    web_attacks_30d = Column(Integer, nullable=True)
    malware_detected_30d = Column(Integer, nullable=True)
    bruteforce_30d = Column(Integer, nullable=True)
    cvss = Column(Numeric, nullable=True)
    epss = Column(Numeric, nullable=True)
    is_kev = Column(Boolean, nullable=True)
    vuln_age_days = Column(Integer, nullable=True)
    created_at = Column(DateTime, default=datetime.utcnow)
    project = relationship("Project", back_populates="assets")
    risks = relationship("Risk", back_populates="asset", cascade="all, delete-orphan")
```

```
class Risk(Base):
    __tablename__ = "risks"

    risk_id = Column(UUID(as_uuid=True), nullable=False)
    version = Column(Integer, nullable=False, default=1)

    project_id = Column(UUID(as_uuid=True), ForeignKey("projects.project_id", ondelete="CASCADE"))
    asset_id = Column(UUID(as_uuid=True), ForeignKey("assets.asset_id", ondelete="CASCADE"))

    risk_type = Column(Text, nullable=False)
    category = Column(Text)
    vulnerability = Column(Text)
    threat = Column(Text)
    probability = Column(Numeric)
    probability_level = Column(Integer)
    impact_level = Column(Integer)
    severity_score = Column(Numeric)
    severity = Column(Text)
    strategy = Column(Text)
    status = Column(Text, default="open")
    confirmed = Column(Boolean, default=False)
    recommendation = Column(Text)
    ocorrencias = Column(Integer, default=1)
    created_at = Column(DateTime, default=datetime.utcnow)

    # Constraints compostas
    __table_args__ = (
        PrimaryKeyConstraint('risk_id', 'version', name="risks_pk"),
    )
    # Relacionamentos
    project = relationship("Project", back_populates="risks")
    asset = relationship("Asset", back_populates="risks")
```

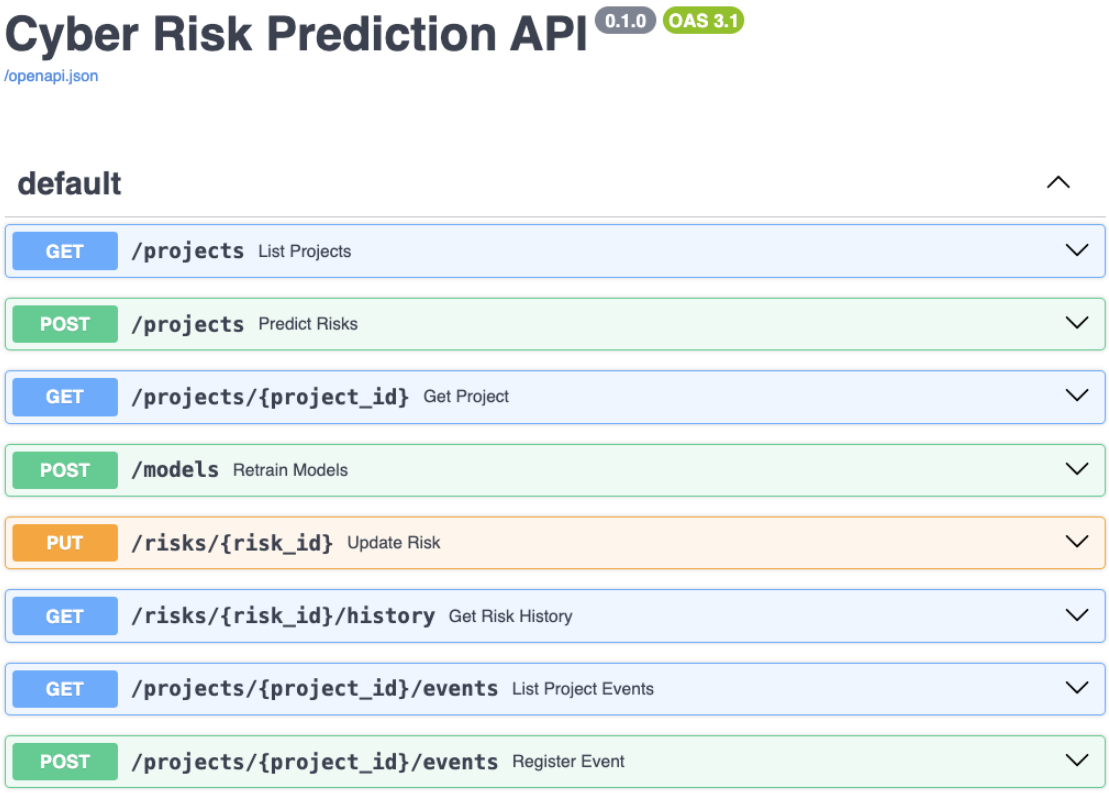
```
class RiskFeedback(Base):
    __tablename__ = "risk_feedback"

    feedback_id = Column(UUID(as_uuid=True), primary_key=True, default=uuid.uuid4)
    risk_id = Column(UUID(as_uuid=True), nullable=False)
    version = Column(Integer, nullable=False)

    project_id = Column(UUID(as_uuid=True), ForeignKey("projects.project_id", ondelete="CASCADE"))
    action_taken = Column(Text)
    comments = Column(Text)
    confirmed = Column(Boolean)
    created_at = Column(DateTime, default=datetime.utcnow)
```

```
class ProjectEvent(Base):
    __tablename__ = "project_events"

    event_id = Column(UUID(as_uuid=True), primary_key=True, default=uuid.uuid4)
    project_id = Column(UUID(as_uuid=True), ForeignKey("projects.project_id", ondelete="CASCADE"))
    asset_id = Column(UUID(as_uuid=True), ForeignKey("assets.asset_id", ondelete="CASCADE"), nullable=True)

    event_type = Column(Text, nullable=False)
    description = Column(Text)
    impact = Column(Text)  # Ex: "mitigate", "increase", "neutral"
    created_at = Column(DateTime, default=datetime.utcnow)

    project = relationship("Project", backref="events")
    asset = relationship("Asset", backref="events")
```

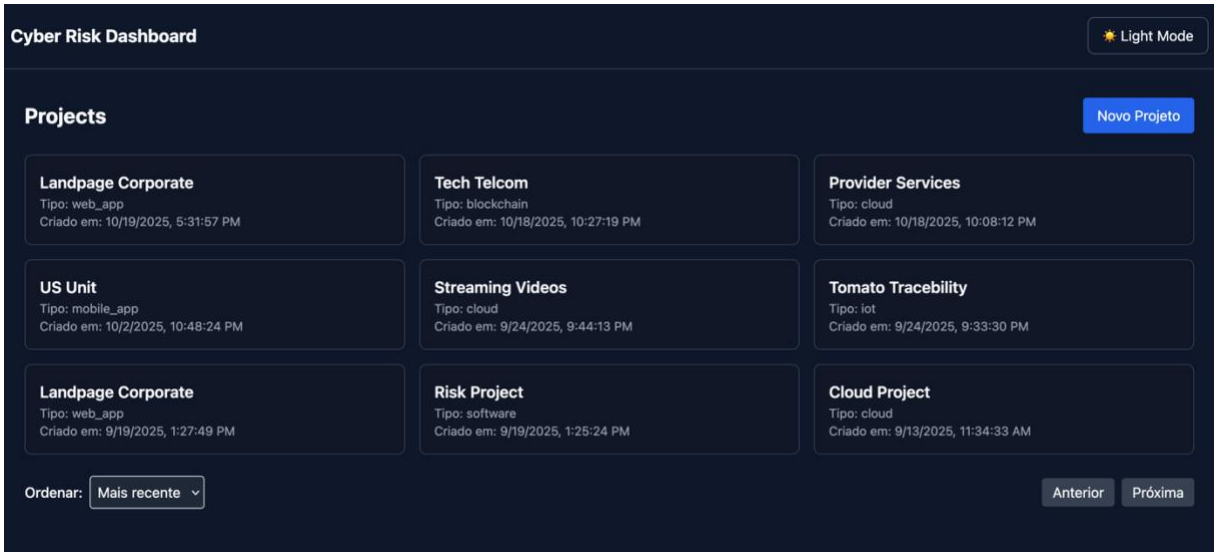
Critério de analise dos riscos no calculo da severidade

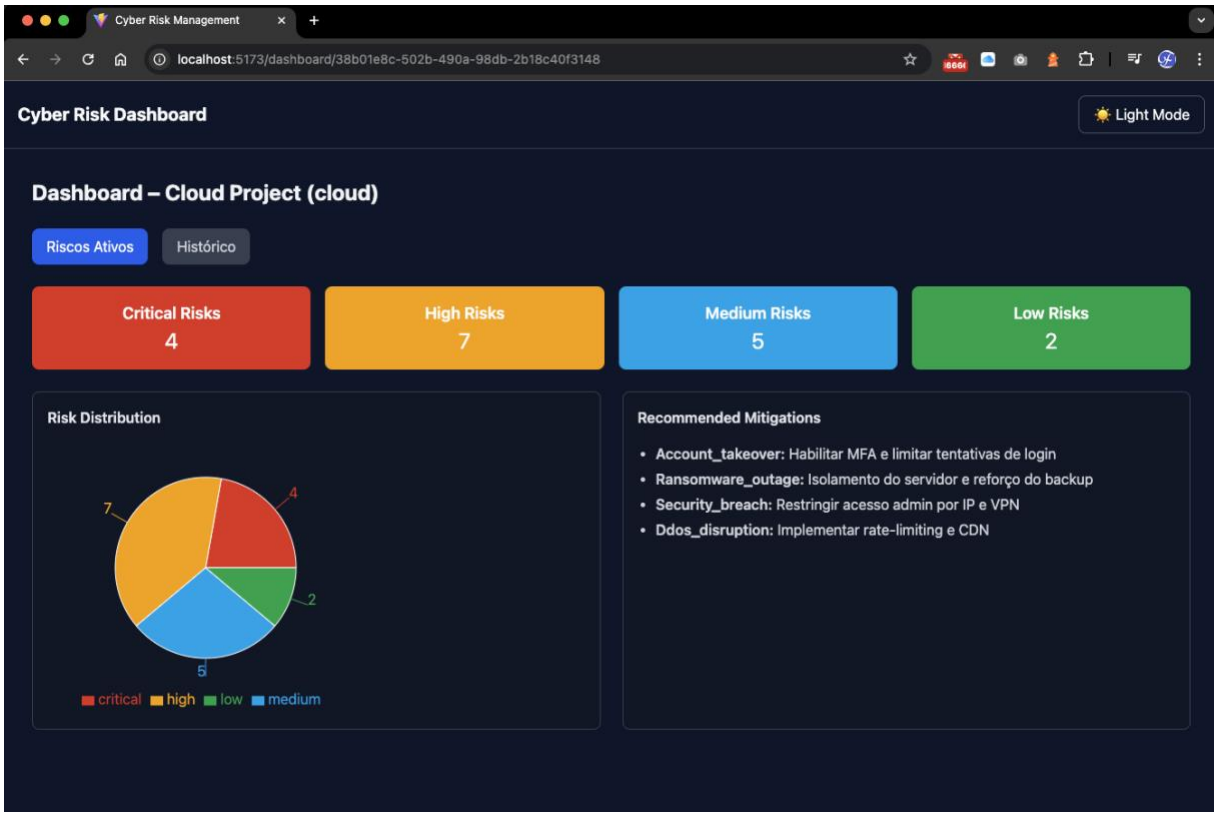
```
def calculate_severity(prob: float, asset) -> dict:
    impact_score = 0
    if get_val(asset, "business_critical", False): impact_score += 2
    if float(get_val(asset, "cvss", 0)) >= 7.0: impact_score += 1
    if get_val(asset, "is_kev", False): impact_score += 1
    if get_val(asset, "alerts_30d", 0) >= 20: impact_score += 1
    if float(get_val(asset, "control_coverage_score", 1)) < 0.4: impact_score += 1
    if get_val(asset, "alerts_30d", 0) > 100: impact_score += 1

    impact_level = min(5, max(1, impact_score))
    severity_score = round((prob * 0.6 + (impact_level / 5) * 0.4) * 10, 2)

    if severity_score >= 9:
        severity = "critical"
    elif severity_score >= 7:
        severity = "high"
    elif severity_score >= 4:
        severity = "medium"
    else:
        severity = "low"

    strategy = ("mitigate" if severity == "critical" or prob >= 0.85 else
                "transfer" if severity == "high" else
                "monitor" if severity == "medium" else "accept")

    return {
        "impact_level": impact_level,
        "severity_score": severity_score,
        "severity": severity,
        "strategy": strategy
    }
```

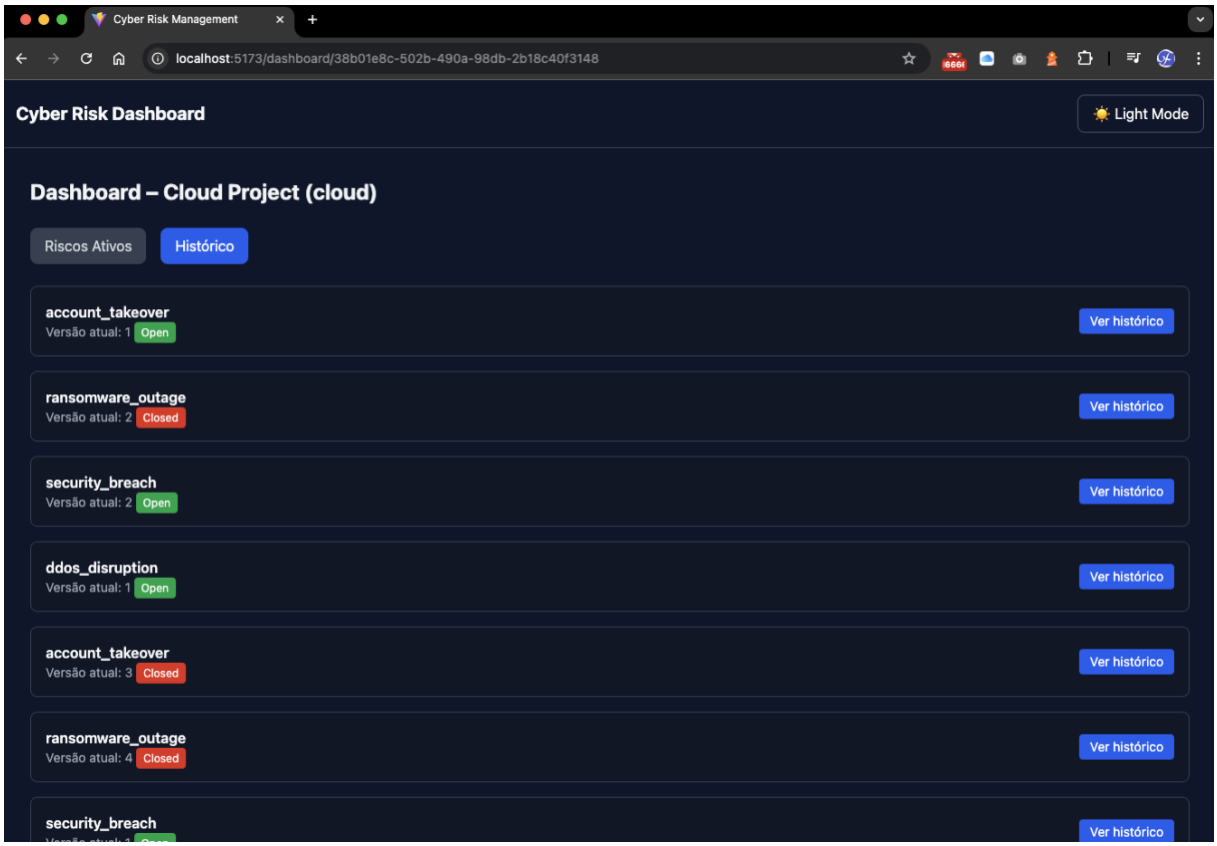
Apêndice C – Endpoints disponíveis do protótipo e GUI

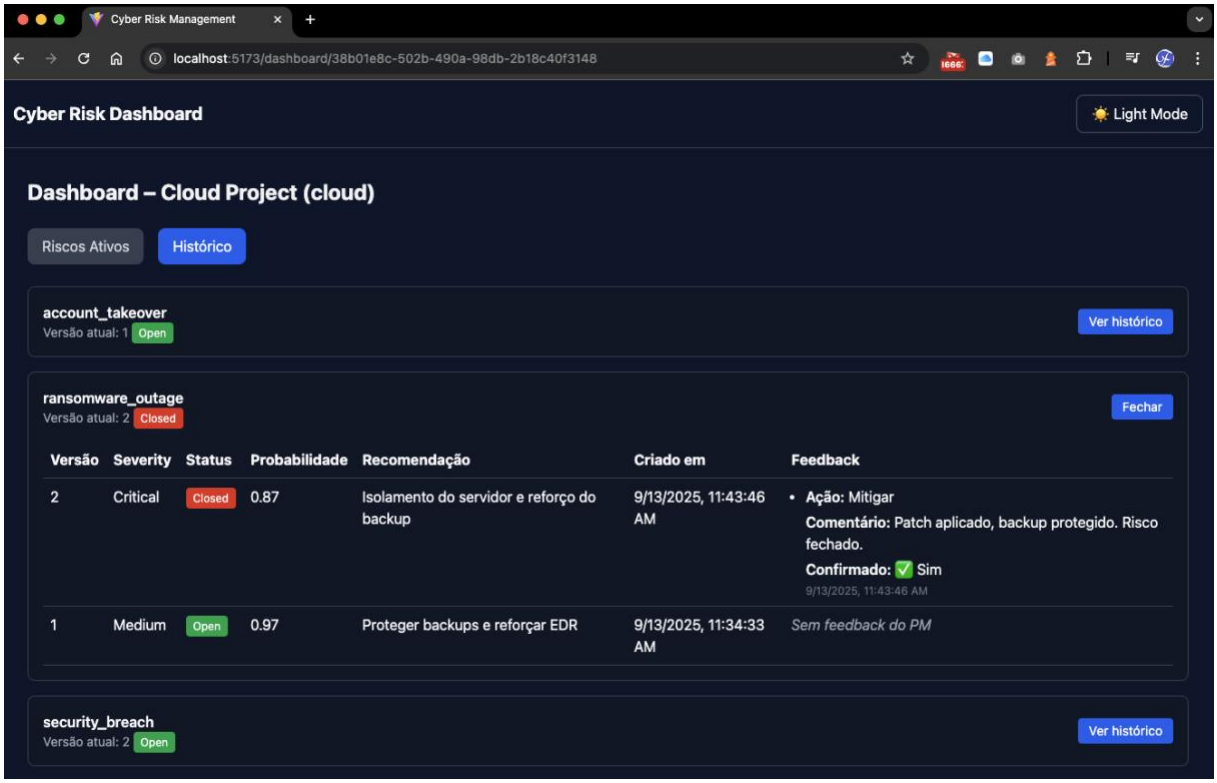


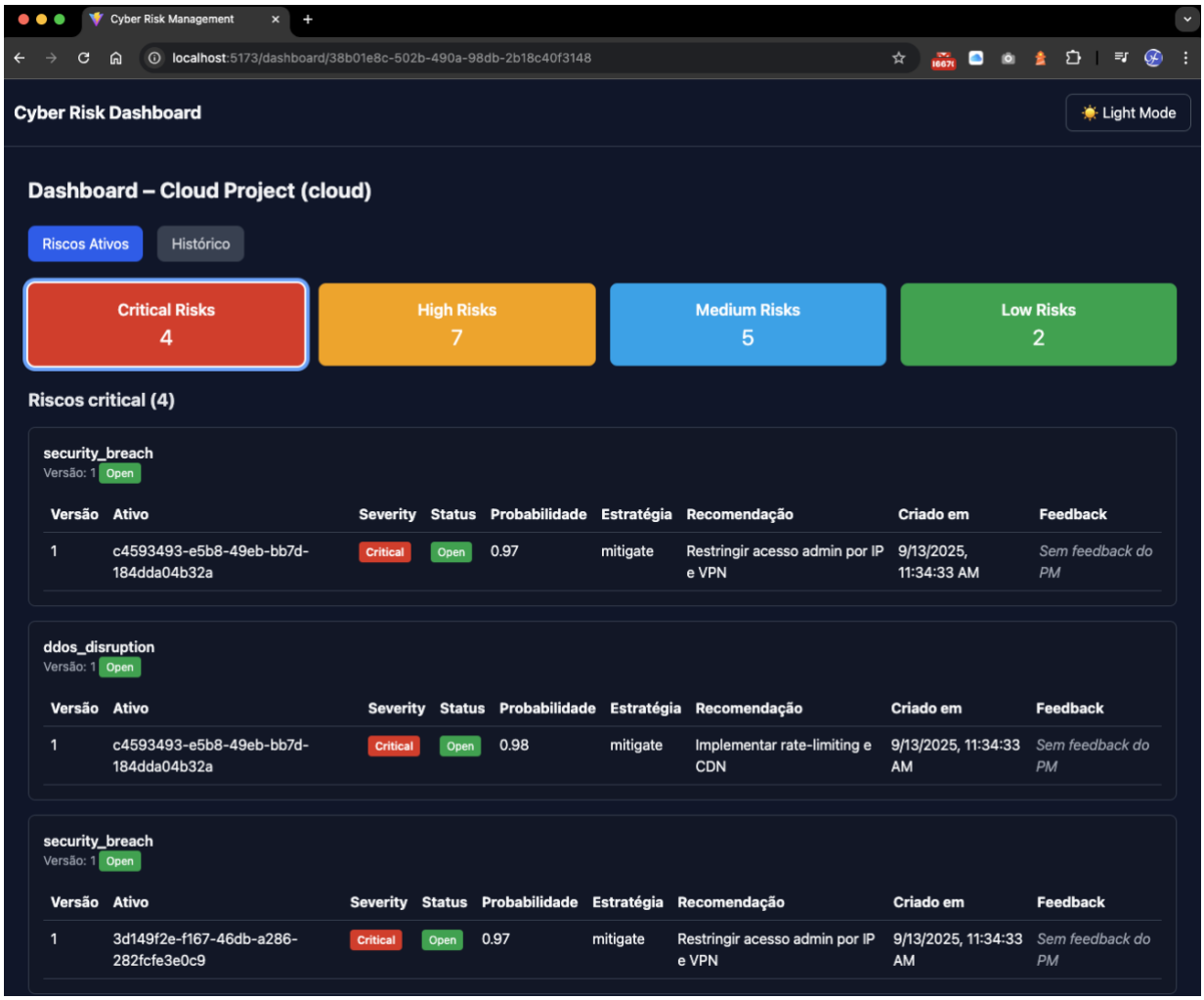
Apêndice D – Frontend para interação com o protótipo





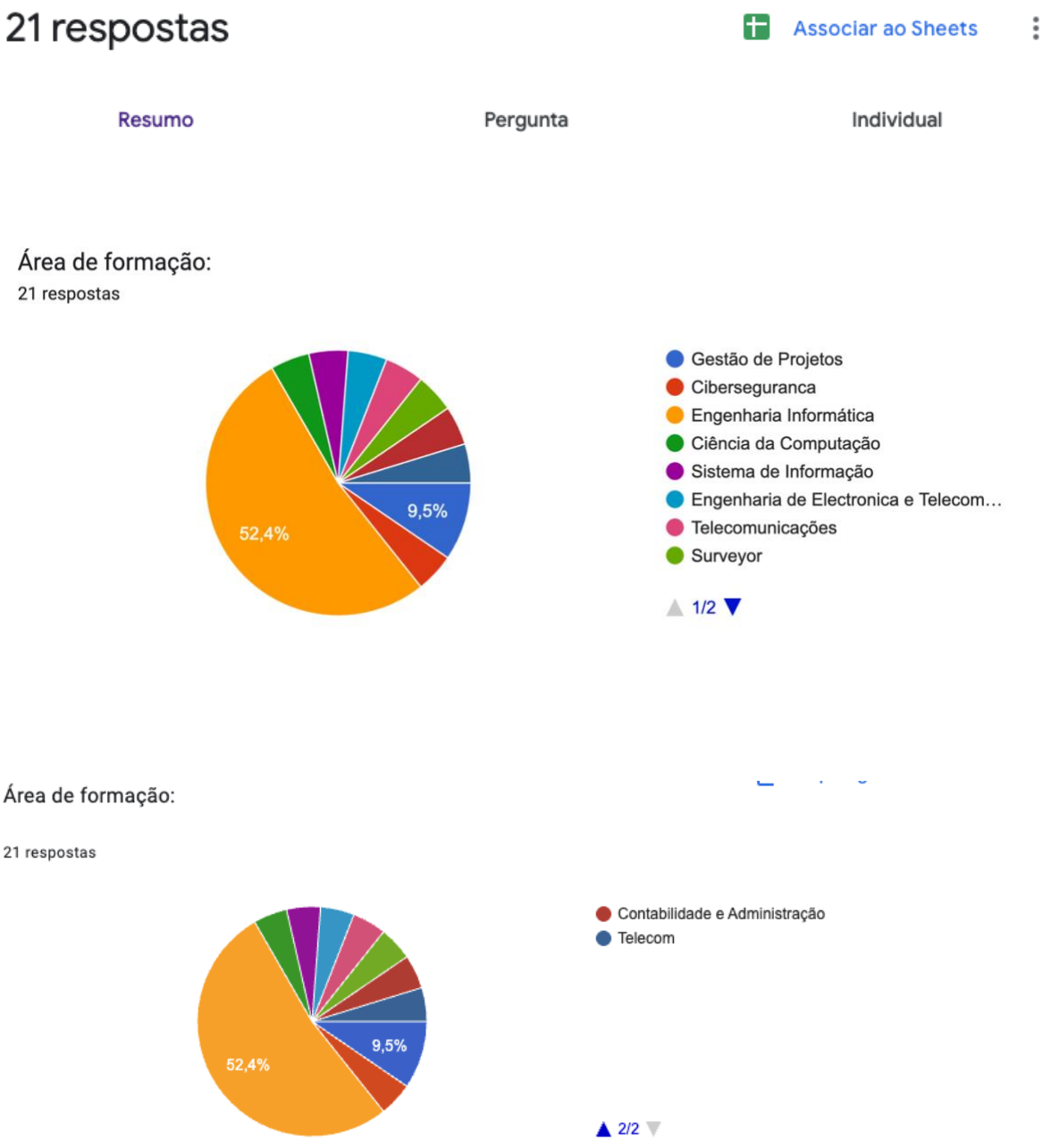

Cyber Risk Management
localhost:5173/dashboard/38b01e8c-502b-490a-98db-2b18c40f3148
Cyber Risk Dashboard
Light Mode
Dashboard – Cloud Project (cloud)
Riscos Ativos
Histórico
account_takeover
Versão atual: 1 Open
Ver histórico
ransomware_outage
Versão atual: 2 Closed
Ver histórico
security_breach
Versão atual: 2 Open
Ver histórico
ddos_disruption
Versão atual: 1 Open
Ver histórico
account_takeover
Versão atual: 3 Closed
Ver histórico
ransomware_outage
Versão atual: 4 Closed
Ver histórico
security_breach
Ver histórico


Cyber Risk Management
localhost:5173/dashboard/38b01e8c-502b-490a-98db-2b18c40f3148
Cyber Risk Dashboard
Light Mode
Dashboard – Cloud Project (cloud)
Riscos Ativos
Histórico
account_takeover
Versão atual: 1 Open
Ver histórico
ransomware_outage
Versão atual: 2 Closed
Fechar
Versão Severity Status Probabilidade Recomendação Criado em Feedback
2 Critical Closed 0.87 Isolamento do servidor e reforço do backup 9/13/2025, 11:43:46 AM
• Ação: Mitigar
Comentário: Patch aplicado, backup protegido. Risco fechado.
Confirmado: ✅ Sim
9/13/2025, 11:43:46 AM
1 Medium Open 0.97 Proteger backups e reforçar EDR 9/13/2025, 11:34:33 AM Sem feedback do PM
security_breach
Versão atual: 2 Open
Ver histórico


Cyber Risk Management
localhost:5173/dashboard/38b01e8c-502b-490a-98db-2b18c40f3148
Cyber Risk Dashboard
Light Mode
Dashboard – Cloud Project (cloud)
Riscos Ativos
Histórico
Critical Risks
4
High Risks
7
Medium Risks
5
Low Risks
2
Riscos critical (4)
security_breach
Versão: 1 Open
Versão Ativo Severity Status Probabilidade Estratégia Recomendação Criado em Feedback
1 c4593493-e5b8-49eb-bb7d-184dda04b32a Critical Open 0.97 mitigate Restringir acesso admin por IP e VPN 9/13/2025, 11:34:33 AM Sem feedback do PM
ddos_disruption
Versão: 1 Open
Versão Ativo Severity Status Probabilidade Estratégia Recomendação Criado em Feedback
1 c4593493-e5b8-49eb-bb7d-184dda04b32a Critical Open 0.98 mitigate Implementar rate-limiting e CDN 9/13/2025, 11:34:33 AM Sem feedback do PM
security_breach
Versão: 1 Open
Versão Ativo Severity Status Probabilidade Estratégia Recomendação Criado em Feedback
1 3d149f2e-f167-46db-a286-282fcfe3e0c9 Critical Open 0.97 mitigate Restringir acesso admin por IP e VPN 9/13/2025, 11:34:33 AM Sem feedback do PM

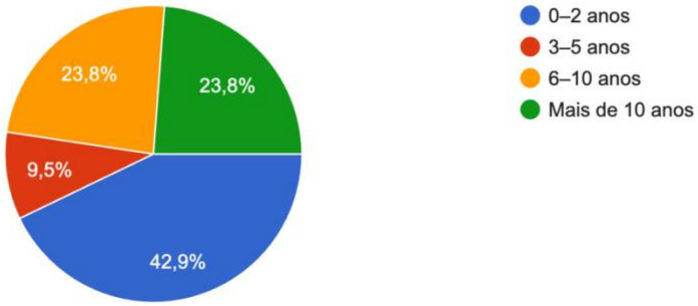
Apêndice E – Inquérito para avaliação do protótipo

Avaliação do Protótipo – Previsão de Riscos de Cibersegurança- Link: https://forms.gle/bcm8WtNaMdQe8s1J6



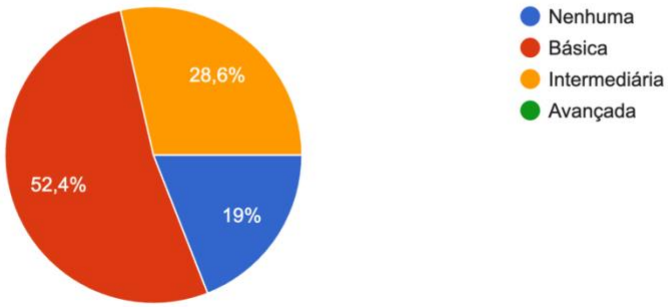
Experiência em Gestão de Projetos

21 respostas



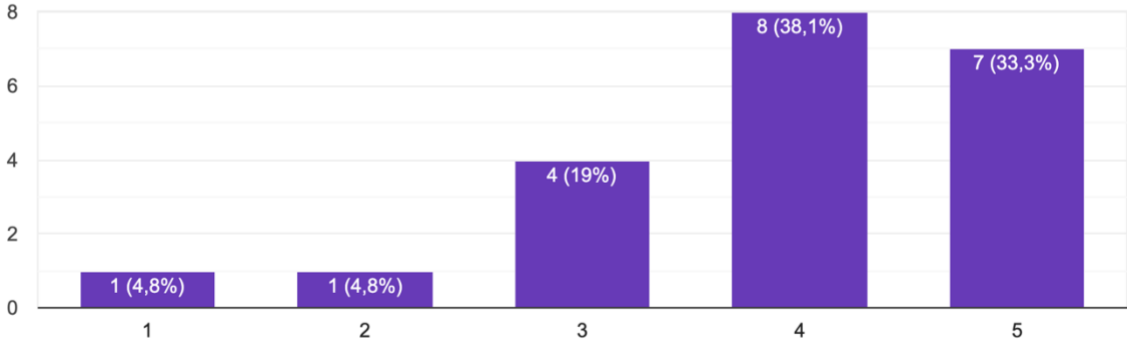
Experiência em Cibersegurança (nível)

21 respostas

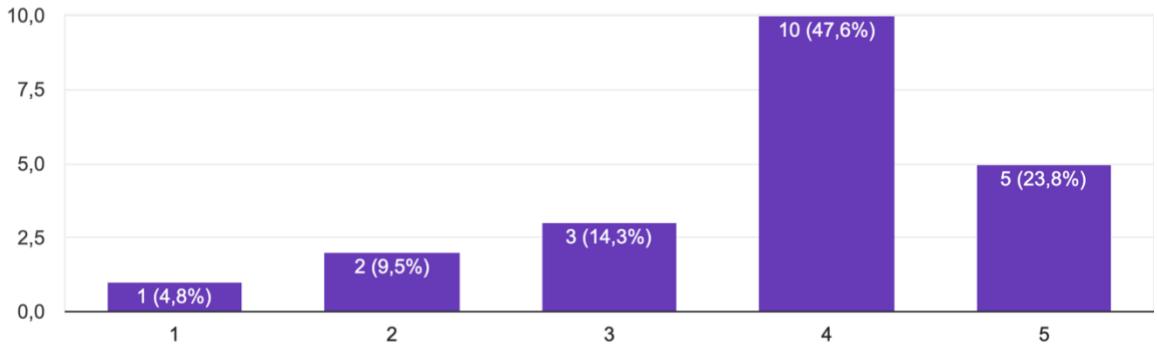


Os riscos de cibersegurança apresentados (ativos, vulnerabilidades, ameaças) são claros e de fácil compreensão?

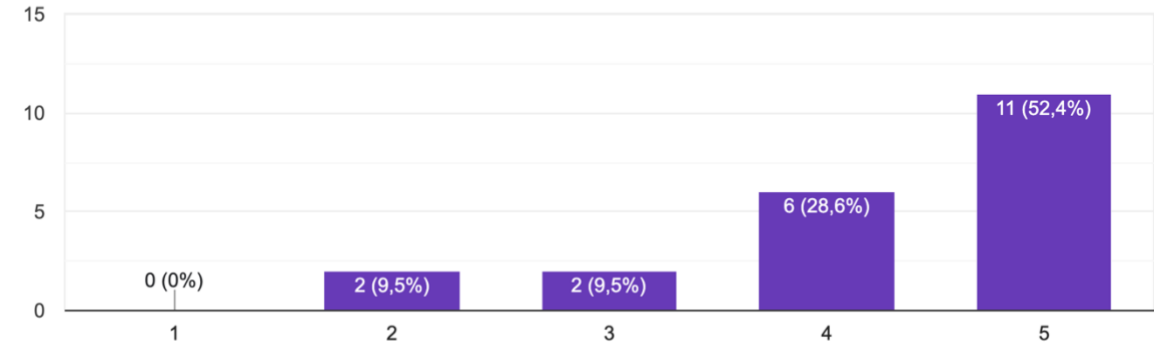
21 respostas



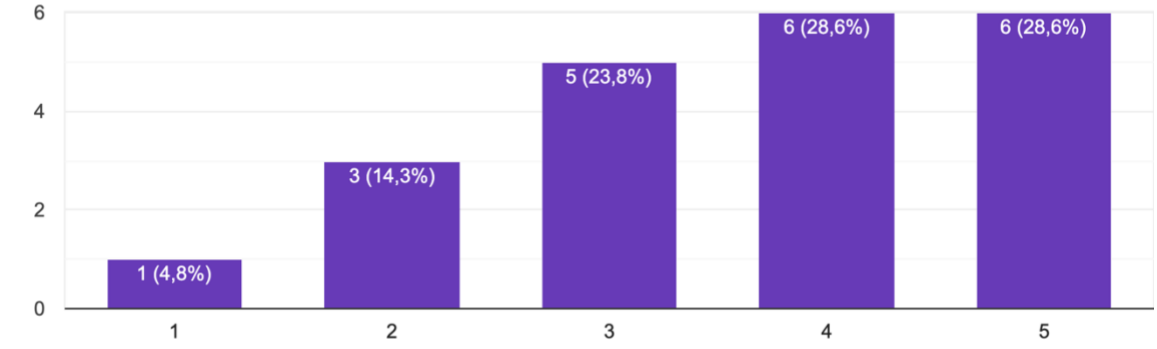
As previsões de riscos refletem cenários realistas de cibersegurança no contexto dos projetos?
21 respostas

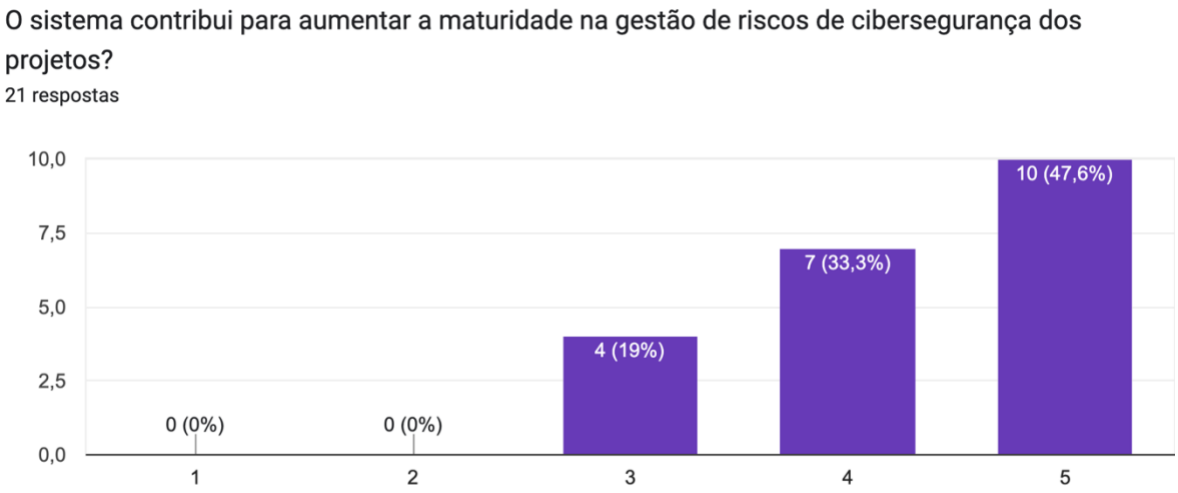


A informação fornecida (probabilidade, impacto, severidade) ajuda a priorizar ações de mitigação em riscos de cibersegurança?
21 respostas



Os riscos previstos são coerentes com a gestão de riscos de TI/cibersegurança?
21 respostas





Que melhorias ou novas funcionalidades relacionadas com a previsão de riscos de cibersegurança sugere?

5 respostas

1. Tecnologias para prevenção, detecção e resposta
2. É um projeto muito pertinente e atual, dada atenção a ter em conta com a segurança dos projetos/sistemas. E como todo protótipo, a tendência é sempre melhorá-lo, e ou atualizá-lo. Muitos parabéns... Palavras de um entusiasta!
3. Gostaria de ver integração com uma plataforma de gestão de projetos.
4. Informações mais clara e simples para todo tipo de pessoas.
5. Integração com sistemas já existentes. Por exemplo, sistemas que analisam bugs e vulnerabilidades em código

Apêndice F - Parâmetros definidos nos algoritmos de Machine Learning

Parâmetros definidos nos algoritmos de Machine Learning durante o processo de treino.

Algoritmo	Parâmetro	Descrição	Função / Observação
XGBoost (XGBClassifier)	n_estimators=150	Número de árvores de decisão a construir	Mais árvores → melhor generalização até certo ponto; 150 oferece equilíbrio entre desempenho e tempo de treino.
	max_depth=5	Profundidade máxima de cada árvore	Controla complexidade e evita overfitting; 5 é uma profundidade moderada.
	learning_rate=0.08	Taxa de aprendizagem (shrinkage)	Reduz o impacto de cada árvore, tornando o modelo mais estável e preciso.
	subsample=0.8	Percentagem de amostras usadas por árvore	Introduz aleatoriedade e evita overfitting.
	colsample_bytree=0.8	Percentagem de features usadas por árvore	Melhora a diversidade entre árvores.
	eval_metric="logloss"	Métrica de avaliação interna	Mede o erro de previsão probabilística (ideal para classificação binária).
	random_state=42	Semente para reprodutibilidade	Garante resultados consistentes entre execuções.
MLPClassifier (Rede Neural)	hidden_layer_sizes=(64, 32, 16)	Estrutura das camadas ocultas	Define 3 camadas densas decrescentes, captando padrões não lineares.
	activation="relu"	Função de ativação	A ReLU (Rectified Linear Unit) é padrão moderna — rápida e evita saturação.
	solver="adam"	Otimizador	Adam combina momentum e adaptação de taxa de aprendizagem — converge bem em dados ruidosos.

	alpha=0.001	Regularização L2	Evita pesos excessivos e overfitting.
	learning_rate_init=0.001	Taxa inicial de aprendizagem	Valor conservador, garante estabilidade.
	max_iter=400	Número máximo de iterações	Dá tempo suficiente para convergir sem overfitting.
	random_state=42	Reprodutibilidade	Fixar semente gera resultados consistentes.
DecisionTree Classifier	max_depth=None	Profundidade máxima da árvore	None permite crescimento completo — ajustado com min_samples para evitar overfitting.
	min_samples_split=4	Mínimo de amostras para dividir um nó	Evita divisões baseadas em amostras muito pequenas.
	min_samples_leaf=3	Mínimo de amostras por folha	Garante folhas representativas e estáveis.
	criterion="gini"	Critério de pureza	O índice de Gini é eficiente para medir impureza de classes.
	random_state=42	Reprodutibilidade	Mantém consistência nos resultados.

Tabela de Síntese Comparativa

Metricas	Modelo	Justificativa
Acurácia Global	XGBoost	Alta consistência em todos os tipos de risco
Precisão	XGBoost	Menor incidência de falsos positivos
Recall	XGBoost	Maior sensibilidade, essencial para riscos de alto impacto
F1-score / F2-score	XGBoost	Equilíbrio ideal entre precisão e recall

ROC AUC	XGBoost	Capacidade de separação entre classes superior
Robustez e Generalização	XGBoost	Alto desempenho médio e menor variabilidade
Interpretabilidade	Decision Tree	Fácil compreensão dos critérios de decisão
Necessidade de Ajustes	MLP	Apresenta desempenho instável e elevado número de erros