1. Introdução

Este relatório tem como objetivo analisar os riscos à proteção de dados pessoais associados à aplicação que realiza a raspagem de informações do Transfermarkt e armazena os dados em arquivos CSV para posterior utilização por um modelo de inteligência artificial (IA), baseado no Ollama com DeepSeek, a fim de responder às perguntas dos usuários. A análise segue a metodologia STRIDE e a modelagem de ameaças de Torr (2005), permitindo a identificação de vulnerabilidades e a proposição de medidas de mitigação.

A arquitetura da aplicação envolve um sistema de raspagem que coleta dados públicos do Transfermarkt, armazena essas informações localmente e as disponibiliza para consulta via IA. Essa abordagem traz desafios relacionados à segurança dos dados processados, armazenamento e controle de acesso, sendo necessário avaliar os riscos envolvidos e definir estratégias para mitigação.

2. Riscos Identificados no Sistema

Para identificar e priorizar os riscos associados à aplicação, utilizamos uma abordagem integrada que combina a modelagem de ameaças proposta por Torr (2005) com a estrutura STRIDE. Essa combinação possibilita a análise sistemática das vulnerabilidades, levando em conta a probabilidade de ocorrência e o impacto potencial de cada ameaça. A seguir, detalhamos as principais categorias de riscos identificadas:

2.1 Falsificação de Identidade (Spoofing)

Aplicando os princípios de Torr (2005), constatamos que a ausência de um mecanismo robusto de autenticação expõe o sistema a ataques de spoofing. Nesse cenário, um invasor pode se passar por um usuário legítimo para obter acesso indevido aos dados raspados do Transfermarkt. Essa vulnerabilidade compromete a confiabilidade dos dados utilizados pela inteligência artificial, impactando diretamente a veracidade das respostas fornecidas. A análise ressalta a necessidade de implementar controles de autenticação (por exemplo, uso de tokens ou autenticação multifator) para reduzir essa ameaça.

2.2 Adulteração de Dados (Tampering)

A integridade dos dados é um pilar fundamental para o funcionamento da aplicação. É crucial identificar pontos onde a manipulação possa ocorrer, uma vez que a transmissão e o armazenamento dos dados em arquivos CSV abrem margem para ataques de tampering. Nesses casos, um agente mal-intencionado pode alterar as informações coletadas, resultando em respostas incorretas pela IA e comprometendo a confiabilidade do serviço. Mecanismos como hashing, assinaturas digitais e controle de versões são recomendados para mitigar esse risco.

2.3 Repúdio (Repudiation)

A falta de um sistema de logs e registros detalhados dificulta a rastreabilidade das ações realizadas dentro da aplicação. Sem a capacidade de auditar e reconstruir o histórico de operações, um usuário mal-intencionado pode negar a autoria de ações indevidas, ampliando o risco de repúdio. Dessa forma, é fundamental registrar todas as interações de forma segura para assegurar a responsabilidade e viabilizar investigações posteriores, mantendo a integridade do controle de acesso.

2.4 Exposição de Informações Confidenciais (Information Disclosure)

Como os dados coletados são todos públicos, o risco de exposição de informações confidenciais é praticamente inexistente. A natureza dos dados raspados do Transfermarkt garante que nenhuma informação sensível ou privada esteja envolvida, eliminando a preocupação com vazamento de dados confidenciais. Dessa forma, mesmo que a agregação e o processamento possam evidenciar padrões ou metadados, esses elementos não comprometem a segurança ou a privacidade dos usuários, pois se referem apenas a informações já de domínio público.

2.5 Negação de Serviço (Denial of Service - DoS)

A continuidade dos serviços oferecidos pela aplicação depende da sua disponibilidade, que pode ser comprometida por ataques de DoS. Esses ataques exploram falhas na infraestrutura ou sobrecarregam o sistema com um volume excessivo de requisições, inviabilizando tanto o processo de raspagem quanto a consulta via IA. Medidas como rate limiting, balanceamento de carga e monitoramento constante do tráfego são essenciais para manter a resiliência da aplicação frente a esse tipo de ameaça.

2.6 Elevação de Privilégio (Elevation of Privilege)

A aplicação não apresenta risco de elevação de privilégio, pois os dados coletados são totalmente públicos e não há níveis diferenciados de acesso que possam ser explorados para obtenção de permissões superiores. Como não há informações sensíveis ou restritas envolvidas no sistema, um eventual acesso indevido não comprometeria a integridade dos dados ou das funcionalidades da aplicação. Dessa forma, essa categoria de ameaça não se aplica ao contexto do projeto.

3. Estratégias de Mitigação dos Riscos

Atualmente, a aplicação não implementa mecanismos específicos para mitigar os riscos identificados. No entanto, existem diversas estratégias que poderiam ser adotadas para fortalecer a segurança e a disponibilidade do sistema. A seguir, apresentamos possíveis medidas que poderiam ser implementadas para minimizar os impactos das ameaças mapeadas.

3.1 Falsificação de Identidade (Spoofing)

A aplicação não possui um sistema de autenticação, permitindo que qualquer usuário acesse os dados raspados sem restrições. Caso fosse necessário controlar o acesso, seria possível adotar mecanismos como autenticação baseada em tokens (JWT) ou autenticação multifator (MFA). Além disso, técnicas como verificação de IPs ou análise de padrões de acesso poderiam ser incorporadas para detectar e bloquear tentativas de acesso indevido.

3.2 Adulteração de Dados (Tampering)

Atualmente, não há mecanismos implementados para garantir a integridade dos dados raspados e armazenados. Para evitar manipulações indesejadas, poderia ser utilizado hashing para verificar a autenticidade dos arquivos CSV, além de assinaturas digitais para garantir que as informações não sejam alteradas após a raspagem. Um controle de versões também poderia ser adotado para permitir a recuperação de dados em caso de alterações não autorizadas.

3.3 Repúdio (Repudiation)

A aplicação não mantém registros detalhados sobre as interações dos usuários e o processamento dos dados. Para possibilitar a rastreabilidade das operações, seria viável implementar um sistema de logs que armazenasse informações como data, hora e ações realizadas, garantindo maior transparência e facilitando auditorias. O uso de logs imutáveis, protegidos por assinaturas digitais, também poderia fortalecer a confiabilidade das informações registradas.

3.4 Exposição de Informações Confidenciais (Information Disclosure)

Como todos os dados processados são públicos, não há risco de exposição de informações confidenciais. Entretanto, para garantir maior organização e evitar interpretações equivocadas sobre a segurança da aplicação, poderia ser estruturado um controle mais rígido sobre os dados armazenados e compartilhados. Além disso, metadados desnecessários poderiam ser removidos antes da publicação dos dados, garantindo um processamento mais eficiente.

3.5 Negação de Serviço (Denial of Service - DoS)

Atualmente, a aplicação não possui mecanismos para prevenir sobrecargas causadas por um grande volume de requisições simultâneas. Caso fosse necessário mitigar esse risco, poderiam ser implementadas técnicas como rate limiting, limitando a frequência de requisições por usuário. Além disso, um balanceador de carga poderia ser utilizado para distribuir as requisições de maneira eficiente, evitando congestionamentos. O monitoramento do tráfego e a detecção de padrões suspeitos também seriam úteis para identificar possíveis tentativas de ataque e prevenir falhas na disponibilidade do serviço.

3.6 Elevação de Privilégio (Elevation of Privilege)

A aplicação não apresenta riscos relacionados à elevação de privilégio, pois todas as informações manipuladas são públicas e não há níveis diferenciados de acesso. No entanto, caso fossem adicionadas funcionalidades com permissões distintas no futuro, seria recomendável adotar o princípio do menor privilégio, garantindo que cada usuário tenha acesso apenas às funções necessárias para seu uso. Auditorias periódicas e revisões de controle de acesso também poderiam ser implementadas para reforçar a segurança do sistema.

4. Conclusão

A análise de impacto à proteção de dados pessoais identificou riscos significativos na aplicação que realiza raspagem de dados do Transfermarkt e os disponibiliza para consulta via IA. A implementação das medidas propostas é essencial para reduzir vulnerabilidades, garantindo a integridade, segurança e conformidade do sistema com boas práticas de proteção de dados.

Monitoramento contínuo, revisão periódica dos mecanismos de segurança e adaptação a novas ameaças devem ser priorizados para assegurar a proteção das informações processadas pela aplicação.