ISSN 1870-4069

Árbol de decisiones en prototipo de sistema de gestión de riesgos para proyectos de software

Axel Omar Alarcón Padilla, Jessie Paulina Guzmán Flores, Juan Jesús Gutiérrez García

Instituto Politécnico Nacional,
Escuela Superior de Cómputo,
México

aalarconp1600@alumno.ipn.mx, jguzmanf@ipn.mx, jgutierrezg@ipn.mx

Resumen. El prototipo se centra en desarrollar una aplicación web que utiliza técnicas de machine learning para la gestión de riesgos en proyectos de software, con un enfoque en la seguridad de la información. Se basa en árboles de decisiones para evaluar y mitigar riesgos, y cuenta con un módulo de plan de seguridad para asignar acciones específicas en respuesta a los riesgos identificados. Además, el prototipo incluye la capacidad de realizar un inventario de activos que clasifica los riesgos según la amenaza, impacto, probabilidad y vulnerabilidad. También implementa un algoritmo para priorizar los riesgos y proporciona informes detallados sobre los riesgos detectados, los planes de seguridad y sus actualizaciones. Adicionalmente, se ofrece una bitácora de seguimiento para registrar los análisis de riesgos y el progreso de los planes de seguridad. Este enfoque integra los estándares del Project Management Institute (PMI) y OWASP, utilizando métricas de análisis de vulnerabilidades, amenazas, probabilidad e impacto para proyectos web, generando una solución integral para identificar, evaluar y priorizar riesgos en función de su probabilidad e impacto.

Palabras clave: Gestión de riesgos, seguridad de la información, machine learning.

Decision Tree in a Risk Management System Prototype for Software Projects

Abstract. The prototype focuses on developing a web application that utilizes machine learning techniques for risk management in software projects, with a particular emphasis on information security. It is based on decision trees to evaluate and mitigate risks and includes a security plan module to assign specific actions in response to identified risks. Additionally, the prototype has the capability to perform an asset inventory that classifies risks according to threat, impact, probability, and vulnerability. It also implements an algorithm to prioritize risks and provides detailed reports on detected risks, security plans, and their updates. Furthermore, a tracking log is available to record risk analyses and

the progress of security plans. This approach integrates the standards of the Project Management Institute (PMI) and OWASP, using metrics for analysing vulnerabilities, threats, probability, and impact for web projects, generating a comprehensive solution to identify, evaluate, and prioritize risks based on their probability and impact.

Keywords: Risk management, information security, machine learning.

1. Introducción

Un software de análisis de riesgos en seguridad de la información es una herramienta esencial para el plan de acción y mantenimiento de cualquier proyecto de software que busque proteger la información, la tecnología y la comunicación, lo cual conlleva a cumplir sus objetivos y asegurar la información de sus usuarios.

A lo largo de la vida escolar, se han realizado proyectos de software, en los cuales se ha observado repetidamente la necesidad de una herramienta robusta de análisis de riesgos en seguridad de la información. Cada proyecto enfrentó desafíos relacionados con la protección de la información y la gestión de amenazas potenciales, lo cual resaltó la importancia de contar con un sistema que facilite la identificación y mitigación de riesgos.

En la experiencia acumulada durante estos proyectos, se evidenció que las soluciones manuales son laboriosas, lo que puede comprometer una correcta evaluación y clasificación de los riesgos. Por lo tanto, se hace imperativo el uso de un software especializado que apoye a los alumnos en estos procesos durante el ciclo de vida de desarrollo de sus proyectos.

La implementación de la herramienta desarrollada se probó con un proyecto de investigación que involucraba una aplicación móvil destinada a apoyar la lectura y comprensión lectora [1], con el fin de garantizar la seguridad de la información dentro del desarrollo de este proyecto. Esta prueba permitió validar la efectividad del software en un entorno real, evaluando su capacidad para identificar, clasificar y gestionar riesgos en las diferentes fases del ciclo de vida del proyecto.

La implementación de un sistema de seguridad de la información requiere el uso de software especializado. Este software utiliza una variedad de técnicas y métodos para identificar, clasificar y evaluar los riesgos asociados a la seguridad de la información en una organización. Este tipo de herramientas es esencial para garantizar la integridad, la confidencialidad y la disponibilidad de los datos almacenados y procesados por los sistemas informáticos. Con su uso, las organizaciones pueden tomar medidas preventivas y correctivas para mitigar los riesgos y proteger su información e infraestructura.

El problema que abordaremos a continuación es la forma en la que los equipos de desarrollo de proyectos realizan su gestión de riesgos, pues cada parte se realiza a mano, lo cual es más tardado y susceptible al error humano, ya que las herramientas gratuitas que se encuentran en línea son plantillas que en el usuario introduce los datos, además de usar parámetros que no son claros y no se adaptan a las necesidades del gestor ni del proyecto.

Toda la información procesada y almacenada por un software está sujeta a amenazas físicas, lógicas y organizacionales. Dicha información puede representar datos importantes de clientes y usuarios, desde un nombre hasta una cuenta bancaria, considerada por una organización como un activo del negocio que tiene un valor, por lo que no se puede escatimar en su protección frente a amenazas. En el proceso de desarrollo de un sistema de software existen riesgos potenciales que amenazan la seguridad de la información, dicha información puede perder confidencialidad, integridad y disponibilidad [2].

El perder confidencialidad indica que la información pase a estar disponible o proporcionada a entidades o procesos no autorizados. El que la información pierda integridad hace que deje de ser precisa y completa, mientras que perder disponibilidad significa que deje de ser accesible y usable en petición de una entidad o proceso autorizados. [3] Un riesgo de seguridad de la información es un efecto de incertidumbre en los objetivos de la seguridad de la información. La causa de los riesgos es un incidente, el cual perjudica al ciclo de vida del desarrollo software en cualquier etapa, por lo que un equipo de desarrollo necesita tener un plan de acción frente a incidentes, compuesto por valoración, identificación, administración y tratamiento de riesgos. [3]

La necesidad de abordar eficazmente los riesgos en seguridad de la información es cada vez más apremiante en un entorno digital en constante evolución. A medida que las organizaciones dependen cada vez más de la tecnología y la interconexión, se vuelven más susceptibles a una variedad de amenazas cibernéticas.

1.1. Trabajos relacionados

La gestión de riesgos en el ámbito específico de los proyectos de software, en lo que respecta a la seguridad de la información, no se ha establecido un enfoque que integre los árboles de decisiones para la evaluación de riesgos. En la mayoría de los casos, las soluciones existentes se basan en conjuntos de reglas fijas y pocas de estas ocupan datos previamente recopilados por la herramienta. En su mayoría las herramientas disponibles no contemplan la aplicación de los estándares del Project Management Institute (PMI) y o del OWASP. Sin embargo, al revisar la literatura actual, se observa que muchos de los enfoques no están alineados claramente con estos estándares, ni detallan de manera específica el uso de árboles de decisiones como herramienta clave para la priorización y gestión de riesgos. Algunas de las herramientas analizadas son:

- Plantillas de matrices y gestión de riesgos en Excel [9]. Estas son las más accesibles y se encuentran en una breve búsqueda en internet, si bien estas son personalizables todas son a la experiencia que tenga el gestor a la hora de categorizar y evaluar un riesgo.
- Pirani [10]. Gestiona procesos, controles, eventos, planes de acción, indicadores y evaluaciones relacionados con el gobierno corporativo, no implementa los estándares del Project Management Institute (PMI) tampoco los de OWASP así como no deja en claro si utiliza arboles de decisiones.
- Aperisoft [11]. Imita cómo funciona el riesgo y la gestión de riesgos en el mundo real. Analiza los riesgos cuantitativamente con un motor de simulación Monte Carlo,

pero no utiliza arboles de decisiones para la evaluación de los riesgos y no implementa los estándares del Project Management Institute (PMI).

2. Módulos

Se ha concebido el desarrollo de un prototipo de aplicación Web de gestión de riesgos altamente especializada, con un enfoque primordial en proyectos de software y tomando como referencia las buenas prácticas de los marcos de trabajo. A diferencia de las alternativas disponibles en el mercado que carecen de su integración y se encuentran diseñadas para organizaciones con una infraestructura de sistemas robusta, el prototipo de aplicación Web está diseñado para ser una solución enfocada en proyectos.

Al centrarnos en la gestión de proyectos de software y aprovechar los principios de la guía de gestión de riesgos del Instituto de Gestión de Proyectos (Project Management Institute, PMI en sus siglas en inglés), en base al PMI se tomó en cuenta el proceso de gestión de riesgos en el ciclo de vida del proyecto. Este enfoque permite integrar la gestión de riesgos en todas las fases del proyecto.

Asimismo, se ha considerado la metodología de valoración de riesgos del Proyecto Abierto de Seguridad de Aplicaciones Web (Open Web Application Security Project, OWASP en sus siglas en inglés). Con el OWASP Risk Rating se utilizó para la evaluación de los riesgos, adaptándolo con un enfoque de inventario de activos. Esta adaptación implica identificar y clasificar los activos del proyecto, lo que nos ayuda a evaluar los riesgos en función de su impacto y probabilidad, nuestro prototipo de aplicación Web brinda una metodología y una estructura para realizar la evaluación de riesgos, lo que nos ayuda a crear planes para mitigar y controlar los riesgos.

El sistema ofrece herramientas para evaluar la probabilidad y el impacto de los riesgos identificados, permitiendo a los usuarios asignar valores numéricos a estos parámetros para calcular la prioridad de los riesgos y establecer estrategias de mitigación.

Mantiene un registro completo de todos los riesgos identificados, incluyendo su descripción y nivel de prioridad.

Los módulos del sistema son los siguientes:

– Módulo de gestión de proyectos y usuarios: En este módulo se manejarán los proyectos de software y las sesiones de los usuarios gestores y participantes, así como el alta de estos en el sistema. El usuario gestor tiene el acceso completo del sistema, junto con todos sus módulos y la capacidad de añadir distintos usuarios participantes que se encarguen de las acciones del módulo del plan de seguridad. Además, se centra en proporcionar a estos usuarios participantes la capacidad de acceder a su sesión personalizada para reportar el progreso de las acciones asignadas en todos los proyectos en los que participan. Los usuarios participantes, que pueden ser supervisores de área, jefes de área, técnicos de área, u otros roles responsables de mantener la integridad y el orden de los reportes de tareas dentro de la organización, tendrán la responsabilidad de reportar el porcentaje de avance de los

planes de acción asignados y proporcionar información detallada sobre dicho avance para el usuario gestor.
- Módulo de inventario de activos: Tendrá formularios para dar de alta, baja o modificar cada activo que tenga el proyecto de software. La sección de activos ahora consta de dos partes: inventario y evaluación. En el inventario de activos el gestor puede visualizar a detalle y editar los activos del proyecto además de poder agregar otros, de acuerdo con las necesidades específicas del proyecto. Los datos de los activos en el inventario se componen de una ficha técnica sin métricas de valoración para el análisis de riesgos. Mientras que, en la sección de evaluación, se le dan valores a las variables de confidencialidad, disponibilidad e integridad de los activos del inventario, dando un valor de sensibilidad del activo.
- Módulo de gestión de riesgos: Este módulo tendrá desglosados cada riesgo identificado con sus activos asociados, y su amenaza, vulnerabilidad, probabilidad de ocurrencia e impacto. Además de la matriz de riesgos con un semáforo colorimétrico en el que se visualizará la valoración de cada riesgo calculada por la aplicación.

2.1. Procesos del módulo de gestión de riesgos

1. Identificación de amenazas: Una vez que se ha establecido un inventario de activos, se procede a identificar las amenazas potenciales que podrían afectar a estos activos. Las amenazas pueden manifestarse en forma de ataques de hackers, malware, ataques de denegación de servicio (DDoS) y otros peligros que podrían comprometer la seguridad de la aplicación web.

2. Evaluación de vulnerabilidades: Este paso implica la evaluación de las vulnerabilidades en la aplicación que podrían ser explotadas por las amenazas identificadas en el paso anterior. Estas vulnerabilidades pueden incluir problemas de seguridad en el código, configuraciones incorrectas en servidores, falta de control de acceso y otras debilidades potenciales.

3. Valoración de riesgos: Implica asignar valores a factores de confidencialidad, integridad y disponibilidad de los activos asociados. Además, se considera la probabilidad e impacto de que ocurra dicho riesgo. Este proceso permite entender la magnitud del riesgo asociado con cada activo [4]. Para calcular la probabilidad, se tienen cuatro factores de amenaza y cuatro de vulnerabilidad. Los factores de amenaza son: Nivel de habilidad, motivación, oportunidad y tamaño.

El valor de la amenaza del riesgo está dado por [4]:

$$\textbf{Amenaza} = \frac{\mathrm{N_H} + M + O + T}{8} \ . \tag{1}$$

donde:

N_H= Nivel de habilidad de la amenaza en un valor del 1 al 9.
M= Motivación de la amenaza en un valor del 1 al 9.
O= Oportunidad de la amenaza en un valor del 1 al 9.

T= Tamaño de la amenaza en un valor del 1 al 9.

Los factores de vulnerabilidad son: Facilidad de descubrimiento, facilidad de explotación, conciencia y detección de intrusiones. El valor de la vulnerabilidad del riesgo está dado por [4]:

$$\textbf{Vulnerabilidad} = \frac{F_D + F_E + C + D_I}{8} \ . \tag{2}$$

donde:

F_D= Facilidad de descubrimiento de la vulnerabilidad en un valor del 1 al 9.
F_E= Facilidad de explotación de la vulnerabilidad en un valor del 1 al 9.
C= Conciencia de la vulnerabilidad en un valor del 1 al 9.
D= Detección de intrusiones en la vulnerabilidad en un valor del 1 al 9.

Con dichos factores, el valor de la probabilidad del riesgo está dado por [4]:

$$\textbf{Probabilidad} = A + V \ . \tag{3}$$

donde:

A= Valor de amenaza derivado de sus factores
V= Valor de vulnerabilidad derivado de sus factores.

Para calcular el impacto, se tienen cuatro factores de impacto empresarial y tres de impacto técnico derivado de los activos asociados al riesgo. Los factores de impacto empresarial son: Daño financiero, daño a la reputación, incumplimiento y violación a la privacidad. El valor del impacto empresarial del riesgo está dado por [4]:

$$\boldsymbol{I_E} = \frac{D_F + D_R + N_C + V_P}{4} \ . \tag{4}$$

donde:

I_E= Valor del impacto empresarial del riesgo.
D_F= Daño financiero a la organización en un valor del 1 al 9.
D_R= Daño a la reputación de la organización en un valor del 1 al 9.
N_C= Incumplimiento de la organización en un valor del 1 al 9.
V_P= Violación de la privacidad de la organización en un valor del 1 al 9.

Los factores del impacto técnico son las variables de confidencialidad, disponibilidad e integridad del activo asociado al riesgo con mayor sensibilidad. El valor del impacto técnico está dado por [4]:

$$\boldsymbol{I_T} = \frac{C_O + D + I_N}{3} . \tag{5}$$

donde:

C_O= Confidencialidad del activo con la mayor sensibilidad de la lista de activos asociados al riesgo en un valor del 1 al 9.

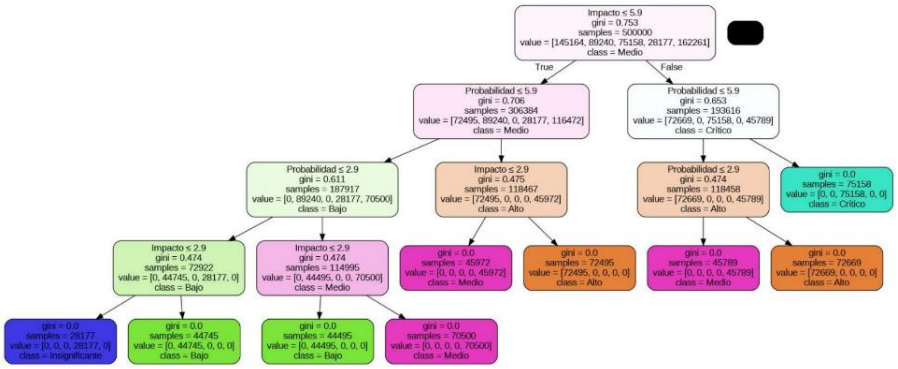

Fig. 1. Árbol de decisiones de la priorización del riesgo.

D= Disponibilidad del activo con la mayor sensibilidad de la lista de activos asociados al riesgo en un valor del 1 al 9.

I_N= Integridad del activo con la mayor sensibilidad de la lista de activos asociados al riesgo en un valor del 1 al 9.

Con dichos factores, el valor del impacto del riesgo está dado por [4]:

$$\textbf{Impacto} = \frac{I_E + I_T}{\mathbf{2}}. \tag{6}$$

4. Análisis del riesgo: Se calcula el riesgo para cada activo multiplicando el impacto y la probabilidad. Este cálculo proporciona una medida cuantitativa de la gravedad del riesgo.

5. Priorización de riesgos: Con la información sobre los riesgos calculados, se priorizan los activos en función de la magnitud de su riesgo. Los activos con riesgos más altos reciben una atención especial en términos de medidas del plan de seguridad. El diagrama de flujo de la priorización representa el árbol de decisiones que solo evalúa las variables de estimación de probabilidad e impacto del riesgo, como lo muestra la Figura 1.

La aplicación de técnicas de machine learning en la metodología de gestión de riesgos aporta una capacidad predictiva y adaptativa al prototipo, permitiendo anticipar posibles amenazas y tomar medidas preventivas de manera proactiva. Al analizar datos históricos y tendencias emergentes, el sistema puede identificar de forma automática escenarios de riesgo potenciales, brindando a los usuarios una herramienta avanzada para la toma de decisiones informadas en materia de seguridad de la información. Asimismo, la estructura de árboles de decisiones facilita la visualización y comprensión de las diferentes opciones de mitigación de riesgos, ofreciendo una guía clara y estructurada para la gestión integral de la seguridad en proyectos de software.

Al visualizar las diferentes herramientas de machine learning se observó que la que más se acomodaba a las necesidades del prototipo era el árbol de decisiones tipo CART

Tabla 1. Comparación de resultados de gestores de riesgo.

Herramienta utilizada	Riesgos detectados	Planes implementados	Riesgos reevaluados
Prototipo	57	81	44
Plantilla	38	63	26

ya que en la metodología de gestión de riesgos evalúa la importancia de diferentes variables y sus impactos en los resultados del proyecto. Esta evaluación precisa ayuda a priorizar los riesgos más críticos y tomar decisiones informadas sobre cómo abordarlos. Esto es una ayuda para gestores de riesgo que no tienen experiencia en el análisis y gestión de riesgos ya que mucha de la dificultad radica en que no han realizado estas actividades con de una forma estructurada. Además, que capacidad predictiva y la estructura jerárquica de decisiones de los árboles de decisiones de este tipo ofrecen anticipar posibles riesgos, esto es especialmente útil en la fase de planificación del proyecto, ya que se pueden tomar medidas preventivas desde ese punto del ciclo de vida del proyecto.

3. Resultados y trabajo a futuro

Se plantearon una serie de pruebas a lo largo del desarrollo e implementación de un proyecto de software, se contrastarían los riesgos detectados, los planes de mitigación implementados y los riesgos reevaluados de la Aplicación móvil para el apoyo a la lectura y comprensión lectora [1] con el prototipo propuesto y la herramienta al alcance del equipo de desarrollo que eran las plantillas de matriz y gestión de riegos en Excel [9] ya que esta aplicación forma parte de un proyecto de investigación y el utilizar otra herramienta saldría del presupuesto del mismo.

Se eligieron el número de riesgos detectados planes implementados y riesgos reevaluados ya que son una forma de medir la diferencia que hay entre una herramienta que ayuda a la gestión a otra que depende totalmente de la experiencia del gestor, por lo que al contabilizar estos parámetros vemos si la herramienta ayuda a gestores de un mismo proyecto con una experiencia similar.

Para la ejecución de las pruebas se tomaron dos integrantes del equipo de desarrollo con una experiencia similar a la gestión de riesgos, los cuales se encargarían de realizar la evaluación y gestión de riesgos con el prototipo o la plantilla y los resultados fueron los siguientes:

Durante las pruebas realizadas en la Aplicación móvil para el apoyo a la lectura y comprensión lectora [1] se resolvió que el prototipo ayudo a la gestión de riesgos de la información e infraestructura durante el desarrollo del proyecto, tomando en cuenta las fases de desarrollo en las cuales se encontraba el mismo, también ayudo a la gestión de

riesgos ya cuando la aplicación de encontraba en un ambiente productivo. Esto nos dice que el prototipo puede ser implementado para que esté al alcance de los compañeros que realizan un proyecto de software para fines escolares, incluyendo proyectos parciales o finales, haciendo énfasis en los trabajos terminales ya que es un requisito hacer un análisis de riesgos por proyecto.

La implementación de una herramienta como la que se propone siendo está el prototipo de gestión de riesgos apoyaría a la población estudiantil a general un análisis de riesgos con una metodología y una estructura para realizar la evaluación de riesgos, lo que nos ayuda a crear planes para mitigar y controlar los riesgos a lo largo del ciclo de vida del desarrollo del proyecto y en algunos casos este proyecto en un ambiente productivo.

Como trabajo a futuro, se plantea la exploración de la implementación de sistemas conscientes del contexto en el prototipo, con el fin de adaptar la gestión de riesgos de manera dinámica según las necesidades y la interacción del usuario con el entorno digital. Se considera que la fusión de datos de fuentes heterogéneas, en línea con la detección de inconsistencias y la adaptación contextual, podría ser un enfoque prometedor para fortalecer la gestión de riesgos en entornos digitales en constante evolución.

Para abordar estos desafíos futuros, se lanzará el prototipo para el uso del mismo dentro de tres unidades de aprendizaje en la Escuela Superior de Computo, siendo estas: Análisis y Diseño, Formulación de Proyectos de TI y Gobierno de TI, teniendo estas 3 relación con el análisis y gestión de riesgos, poniendo a disposición la herramienta para los alumnos y profesores, probando en diferentes escenarios y contextos la herramienta y poblando la base de datos.

También se pondrá a disposición de los alumnos que estén realizando su Trabajo Terminal tanto la primer y segunda parte, esto para observar cómo se realiza un análisis de riesgos en proyectos de software, para la primera parte de los Trabajos Terminales, y el seguimiento de los mismos a lo largo del desarrollo del proyecto.

También se propone la investigación y aplicación de métodos de fusión de datos de fuentes heterogéneas en el contexto de la gestión de riesgos en proyectos de software. Se busca mantener la consistencia de los datos y mejorar la capacidad predictiva del sistema mediante la integración de información contextual [8] diversa.

Se espera que esta línea de trabajo contribuya a fortalecer la seguridad de la información en proyectos de software y a proporcionar herramientas más avanzadas y adaptativas para la gestión de riesgos en un entorno digital cada vez más complejo y dinámico.

Agradecimientos. Los resultados de este trabajo se desarrollaron en el marco del proyecto de investigación: Aplicación móvil para el apoyo docente en los procesos enseñanza - aprendizaje de la lectura mediante técnicas de gamificación y machine learning con número de registro asignado por el SIP: 20242235, desarrollado en Instituto Politécnico Nacional.

Referencias

1. Rodríguez, O.E., Huerta, M., Reyes, E.D.: Trabajo Terminal: Aplicación móvil para el apoyo a la lectura y la comprensión lectora. Escuela Superior de Cómputo, Instituto Politécnico Nacional (2024)
2. Ieeeauthorcenter: Information security, cybersecurity and privacy protection, Information security management systems, Requirements. In: ISO/IEC 27001:2022. https://ieeeauthorcenter.ieee.org/wp-content/uploads/IEEE-Reference-Guide.pdf (2022)
3. Standards: Information technology, Security techniques – Information security management systems, In: Overview and vocabulary ISO/IEC 27000:2018. https://standards.iso.org/ittf/PubliclyAvailableStandards/index.html (2018)
4. OWASP: Risk Rating Methodology. https://owasp.org/www-community/OWASP_ Risk_ Rating_Methodology (2022)
5. Project Management Institute: Practices Standard for Project Risk Management. Project Management Institute, Inc., USA (2009)
6. Information security: Cybersecurity and privacy protection: Information security controls. ISO/IEC 27002:2022. International Organization for Standardization (2022)
7. Information Technology: Security techniques -Information security risk management. ISO/IEC 27005:2018. International Organization for Standardization (2018)
8. Julio, M., Guillermo, M.C., Edgard, B.G.: Método de fusión de datos de fuentes heterogéneas para mantener la consistencia de datos. Research in Computing Science Issue, 139, pp. 33–46 (2017)
9. Smartsheet: Plantillas de matriz y gestión de riegos en Excel. https://es.smartsheet.com/all-risk-assessment-matrix-templates-you-need (2023)
10. Pirani: https://www.piranirisk.com/es/ (2023)
11. Aperitisoft: https://info.gartnerdigitalmarkets.com/rpm3solutions-gdm-lp?category=integrated-risk-management&utm_source=Capterra (2023)
12. NVIDIA: Pandas Python. https://www.nvidia.com/en-us/glossary/pandas-python/ (2023)
13. Pandas Documentation: https://pandas.pydata.org/docs/user_guide/dsintro.html (2023)
14. Fine Tuning LLMs: DataCamp. https://www.datacamp.com/tutorial/fine-tuning-large-language-models (2024)
15. Vectorization techniques: https://neptune.ai/blog/vectorization-techniques-in-nlp-guide (2023)
16. Understanding TF-IDF: Geeks for Geeks. https://www.geeksforgeeks.org/understanding-tf-idf-term-frequency-inverse-document-frequency/ (2023)
17. TF-IDF: LearnDataSci. https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/ (2024)
18. Classification in Machine Learning: DataCamp. https://www.datacamp.com/blog/classification-machine-learning (2024)
19. Random Forest Classifier using Scikit-learn: Geeks for Geeks. https://www.geeksforgeeks.org/random-forest-classifier-using-scikit-learn/ (2024)