

CENTRO UNIVERSITÁRIO UNIDADE DE ENSINO SUPERIOR DOM BOSCO
CURSO ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

PAULO JOSÉ FERREIRA MARQUES

**A IMPORTÂNCIA DA MINERAÇÃO DE DADOS NA CRIAÇÃO DE INDICADORES
PREDITIVOS PARA O GERENCIAMENTO DE RISCOS CORPORATIVOS**

São Luís

2022

PAULO JOSÉ FERREIRA MARQUES

**A IMPORTÂNCIA DA MINERAÇÃO DE DADOS NA CRIAÇÃO DE INDICADORES
PREDITIVOS PARA O GERENCIAMENTO DE RISCOS CORPORATIVOS**

Monografia apresentada ao Curso de Análise e Desenvolvimento de Sistemas do Centro Universitário Unidade de Ensino Superior Dom Bosco como requisito parcial para obtenção do grau de Tecnólogo em Análise e Desenvolvimento de Sistemas.
Orientador: Prof. Dr. Giovanni Lucca França Da Silva.
Coorientador: Prof. Dr. Darlan Bruno Pontes Quintanilha.

São Luís

2022

Dados Internacionais de Catalogação na Publicação (CIP)
Centro Universitário – UNDB / Biblioteca

Marques, Paulo José Ferreira

A importância da mineração de dados na criação de indicadores preditivos para gerenciamento de riscos corporativos. / Paulo José Ferreira Marques. __ São Luís, 2022.

31 f.

Orientador: Prof. Dr. Giovanni Lucca França Da Silva.

Monografia (Graduação em Sistemas de Informação) - Curso Tecnológico de Análise e Desenvolvimento de Sistemas - Centro Universitário Unidade de Ensino Superior Dom Bosco - UNDB, 2022.

1. Gerenciamento de Riscos Corporativos. 2. Indicador preditivo. 3. Mineração de Dados. 4. Modelos preditivos. I. Título.

CDU 004.65

PAULO JOSÉ FERREIRA MARQUES

**A IMPORTÂNCIA DA MINERAÇÃO DE DADOS NA CRIAÇÃO DE INDICADORES
PREDITIVOS PARA O GERENCIAMENTO DE RISCOS CORPORATIVOS**

Monografia apresentada ao Curso de Análise e Desenvolvimento de Sistemas do Centro Universitário Unidade de Ensino Superior Dom Bosco como requisito parcial para obtenção do grau de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Aprovada em: ____/____/____.

BANCA EXAMINADORA:

Prof. Dr. Giovanni Lucca França Da Silva (Orientador)

Doutor em Engenharia Elétrica

Centro Universitário Unidade de Ensino Superior Dom Bosco (UNDB)

Prof. Dr. Darlan Bruno Pontes Quintanilha (Coorientador)

Doutor em Engenharia de Eletricidade

Universidade Federal do Maranhão (UFMA)

Prof. Me. Rodrigo Monteiro de Lima

Mestre em Ciência da Computação

Centro Universitário Unidade de Ensino Superior Dom Bosco (UNDB)

Prof. Igor Luciano Cavalcanti Lima

Especialista em Análise de Dados com BI e Big Data

Centro Universitário Unidade de Ensino Superior Dom Bosco (UNDB)

Dedico à minha mãe, minha
irmã e meu sobrinho.

AGRADECIMENTOS

Agradeço à minha mãe, Magnólia Ferreira, por sempre ter me incentivado na busca por estudos e o conhecimento.

À minha irmã e parceira de vida, Amanda Marques, que me deu o melhor presente da vida, meu sobrinho Gabriel.

À minha namorada, Renata Ribeiro, por todo o amor, apoio incondicional, compreensão e companheirismo durante toda essa jornada.

Aos amigos de turma, Yuri Ageme e Paulo Giovanni, por todos os bons momentos e ensinamentos compartilhados durante o curso.

Ao meu orientador, Giovanni Silva e ao coorientador Darlan Quintanilha, pela compreensão, prestatividade e dedicação durante todo o período.

Ao professor de TCC Maurício Costa, pela disponibilidade e paciência infinita com seus alunos, mostrando sempre vontade em ensinar e amor pelo que faz.

E a todos meus amigos e família, que são parte fundamental em minha vida.

“Tudo na vida é gerenciamento de risco, não sua eliminação.”
(WRISTON).

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1 – Diferença entre KRI e KPI | 17 |
| Figura 2 – Linha do tempo de um risco. | 18 |
| Figura 3 – Etapas do processo de KDD. | 20 |
| Figura 4 – Multidisciplinaridade da Mineração de Dados. | 20 |
| Figura 5 – Fluxo da criação de um modelo preditivo | 21 |
| Figura 6 – Funcionamento do classificador KNN..... | 22 |
| Figura 7 – Árvore de decisão simples para um carro. | 23 |
| Figura 8 – Árvore de decisão para taxa de juros | 24 |
| Figura 9 – Gráfico de regressão linear. | 25 |
| Figura 10 – Erros comuns na etapa de treinamento de dados | 27 |
| Figura 11 – Modelo de KRI..... | 28 |
| Figura 12 – Modelo de reporte de um KRI. | 29 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|------|---|
| GRC | Gerenciamento de Riscos Corporativos |
| KPI | <i>Key Performance Indicator</i> |
| KRI | <i>Key Risk Indicator</i> |
| KDD | <i>Knowledge Discovery in Databases</i> |
| NBR | Norma Brasileira |
| UNDB | Unidade de Ensino Superior Dom Bosco |
| TCC | Trabalho de Conclusão de Curso |

A IMPORTÂNCIA DA MINERAÇÃO DE DADOS NA CRIAÇÃO DE INDICADORES PREDITIVOS PARA O GERENCIAMENTO DE RISCOS CORPORATIVOS

THE VALUE OF DATA MINING IN CREATING PREDICTIVE INDICATORS FOR THE ENTERPRISE RISK MANAGEMENT

Paulo José Ferreira Marques¹

Giovanni Lucca Franca da Silva²

Darlan Bruno Pontes Quintanilha³

RESUMO

O presente trabalho aborda a dificuldade que as organizações encontram na criação de indicadores preditivos, que antecipam e monitoram as causas dos riscos, e a importância da Mineração de Dados nesse processo, através de técnicas, métodos estatísticos e algoritmos. Para tanto, adotou-se uma pesquisa introdutória de finalidade básica, de objetivo exploratório, abordagem qualitativa, mediante revisão de literatura especializada sobre o tema. Diante disso, apresentou-se o fluxo mais indicado para a criação de modelos preditivos e as principais técnicas da Mineração de Dados relacionadas ao tema Gerenciamento de Riscos Corporativos. Finalizou-se com sugestões de reporte e monitoramento dos indicadores preditivos aos gestores e à alta administração.

Palavras-chave: Gerenciamento de Riscos Corporativos. Indicador preditivo.

Indicador Chave de Risco. Mineração de Dados. Modelos preditivos.

ABSTRACT

This work addresses the struggle faced by organizations in creating predictive indicators, which anticipate and monitor the causes of risks, and the relevance of Data Mining in this process, through techniques, statistical methods and algorithms. Therefore, an introductory survey with a basic purpose was carried out, whose exploratory study with a qualitative approach and formulated through a bibliographic and documentary survey on the subject. In view of this, the most suitable flow for the creation of predictive models and the main techniques of Data Mining relating to the

¹ Graduando do 5º Período do Curso de Análise e Desenvolvimento de Sistemas do Centro Universitário Unidade de Ensino Superior Dom Bosco - UNDB. E-mail: paulo.joseph@hotmail.com.

² Professor Docente do Curso de Análise e Desenvolvimento de Sistemas do Centro Universitário Unidade de Ensino Superior Dom Bosco - UNDB. E-mail: giovanni.silva@undb.edu.br.

³ Professor Docente do curso de Engenharia Elétrica da Universidade Federal do Maranhão - UFMA. E-mail: darlan.quintanilha@ufma.br.

Enterprise Risk Management Theme was presented. It ended with suggestions for reporting and monitoring predictive indicators to managers and senior management.

Keywords: Enterprise Risk Management. Predictive indicator. Key Risk Indicator. Data Mining. Predictive models.

SUMÁRIO

| | |
|---|----|
| 1 INTRODUÇÃO | 14 |
| 2 REFERENCIAL TEÓRICO | 15 |
| 2.1 Gerenciamento de Riscos Corporativos | 15 |
| 2.1.1 Indicadores Chave de Performance | 16 |
| 2.1.1 Indicadores Chave de Risco..... | 16 |
| 2.2 Mineração de Dados | 19 |
| 3 PROCESSO DE PREDIÇÃO DE DADOS | 21 |
| 3.1 Ferramentas de predição de dados | 21 |
| 3.1.1 Classificador KNN | 22 |
| 3.1.2 Árvores de decisão..... | 23 |
| 3.1.3 Regressão Linear | 24 |
| 4 RESULTADOS E DISCUSSÃO | 25 |
| 5 CONSIDERAÇÕES FINAIS | 29 |
| REFERÊNCIAS | 31 |

1 INTRODUÇÃO

No mundo corporativo, as organizações têm como principal objetivo agregar valor e gerar lucros para as partes interessadas, e uma forma de identificar a eficiência em relação aos objetivos estratégicos propostos é estipular metas e analisar periodicamente os resultados obtidos, através de indicadores de performance. Apesar de serem relevantes e difundidos mundo afora, eles ilustram apenas um cenário que já ocorreu, sendo chamados de “indicadores atrasados” (NICHOLSON; BAKER, 2013).

Em um ambiente cada vez mais competitivo e cheio de incertezas, faz-se necessário encontrar maneiras de sobressair perante os concorrentes, seja maximizando os lucros ou evitando surpresas negativas que possam impactar nos resultados. Para que isso ocorra, deve-se construir indicadores preditivos, capazes de antecipar cenários, assegurando uma maior assertividade na tomada de decisões.

Ainda segundo Nicholson & Baker (2013), o desafio na construção de indicadores preditivos consiste em identificar a melhor maneira de estabelecer correlações entre os dados analisados e realizar uma projeção assertiva que assegure a capacidade de tomar melhores decisões, de forma antecipada. Existe uma dificuldade em encontrar material disponível sobre o assunto em língua portuguesa e em aplicar o conhecimento de ferramentas estatísticas ao contexto dos indicadores preditivos.

Este estudo tem como objetivo geral auxiliar as organizações e profissionais que buscam criar indicadores preditivos e não possuem conhecimento das técnicas estatísticas e ferramentas da mineração de dados suficientes para realizar tal atividade.

Como objetivos específicos estão: definir conceitualmente como funcionam os indicadores de performance e indicadores de risco, o que é a mineração de dados e quais técnicas e ferramentas podem ser utilizadas na construção de indicadores preditivos.

O presente trabalho foi realizado através de uma pesquisa de natureza básica, com objetivo de caráter exploratório, utilizando uma abordagem qualitativa da literatura escolhida e o procedimento escolhido foi a pesquisa bibliográfica.

Na segunda seção, referencial teórico, objetiva-se apresentar o conceito de Gerenciamento de Riscos Corporativos, sua função nas organizações e principais

diferenças entre indicadores de performance e indicadores de risco, através de literatura especializada sobre o assunto. Respeitando o escopo do trabalho, há um foco maior nos indicadores de risco, capazes de antecipar acontecimentos futuros e minimizar os impactos negativos dos riscos. Apresenta-se também a definição de mineração de dados, áreas de conhecimento que servem de base para a criação de análises e projeções assertivas

Na terceira seção, são apresentados exemplos de ferramentas de mineração de dados focadas na predição de dados, identificando padrões em prol de projeções assertivas. Serão apresentadas as principais técnicas que possuem relação com o tema proposto e suas aplicações para realizar correlações entre dados analisados e construção de indicadores preditivos, como um Indicador Chave de Risco (*traduzido do inglês Key Risk Indicator - KRI*).

Na quarta seção, é apresentado o resultado do trabalho e uma discussão a respeito dos tópicos abordados, etapas do processo de criação dos modelos preditivos, bem como os possíveis problemas e erros resultantes desse ciclo. Apresenta-se também, um modelo de indicador chave de risco e reporte para a gestão.

Na quinta seção, encontram-se as considerações finais do trabalho, a relevância do Gerenciamento de Riscos Corporativos na geração de valor para as organizações e a importância da Mineração de Dados em projeções mais assertivas que alimentam os indicadores preditivos.

2 REFERENCIAL TEÓRICO

Esta seção visa abordar os principais conceitos que servem de base para o desenvolvimento deste trabalho. São apresentadas informações sobre Gerenciamento de Riscos Corporativos, Indicadores de Performance, Indicadores de Risco e a importância Mineração de Dados e emprego de suas técnicas no processo de criação de indicadores preditivos.

2.1 Gerenciamento de Riscos Corporativos

O conceito de Gerenciamento de Riscos Corporativos (GRC) pode ser encontrado no COSO ERM (Enterprise Risk Management) – Gerenciamento de Riscos Corporativos, documento que norteia a atividade de GRC:

Uma vez definida a estratégia, o gerenciamento de riscos corporativos constitui um meio eficaz para a administração cumprir seu papel, uma vez que a organização demonstra estar atenta aos riscos que podem afetar sua estratégia e apta a gerenciá-los. A utilização do gerenciamento de riscos corporativos gera confiança e proporciona segurança aos *stakeholders*, especialmente no contexto atual, que exige um nível de vigilância jamais visto sobre como o risco é tratado e gerenciado. (NICHOLSON, 2017, p. 2)

Em outras palavras, o GRC representa a autoridade que lida com as incertezas e supervisiona constantemente o estado atual das organizações frente a seus objetivos estratégicos, avaliando os cenários interno e externo e verificando o nível de preparo frente aos riscos oriundos do negócio, propondo a criação de novos controles e solicitando respostas perante os riscos de não atingimento dos objetivos.

2.1.1 Indicadores Chave de Performance

Os Indicadores Chave de Performance, conhecidos como *Key Performance Indicators* (KPI) foram criados com intuito de verificar se os resultados estratégicos estão sendo alcançados. Tem como ponto de partida o planejamento estratégico da organização e os objetivos estratégicos distribuídos em metas definidas de acordo com as prioridades da entidade (FALCONI, 2013).

Geralmente são tratados como as próprias metas da organização, acompanhados periodicamente e apresentados em reuniões de resultados, podendo ser atrelados a uma bonificação para os empregados em caso de sucesso. Seu funcionamento é simples e tem como condição para o atingimento apenas um valor previsto, que pode ser alcançado ou não.

2.1.2 Indicadores Chave de Risco

Já os Indicadores Chave de Risco, conhecidos como *Key Risk Indicators* (KRI) são menos conhecidos e possuem um funcionamento diferente dos *KPI's*, conforme a Figura 1:

Figura 1 – Diferença entre KPI e KRI

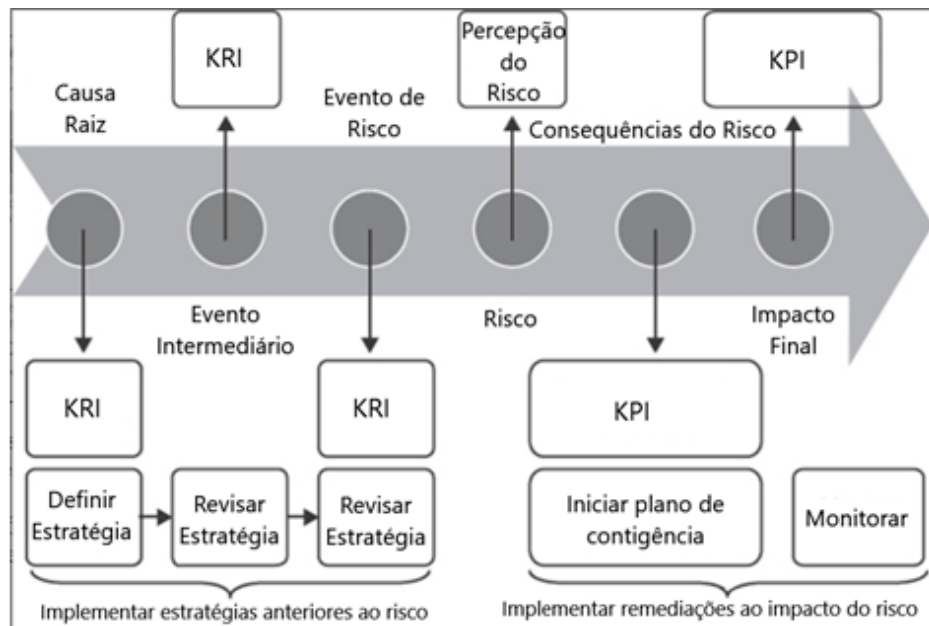


Fonte: NICHOLSON & BAKER (2013, p. 230)

Para Nicholson (2012), enquanto os *KPI's* medem eventos que já ocorreram e ilustram resultados reais, os *KRI's* devem ser criados com base nas causas raízes que podem levar à materialização de um risco considerado como crítico para organização, tendo como base seu planejamento estratégico. No exemplo da Figura 1, percebe-se que a queda no número das vendas é percebida através da utilização de um *KPI* que mede as vendas, porém, o evento que realmente resulta no decréscimo das vendas é a diminuição nas compras, por parte dos clientes. Continuando o raciocínio, chega-se à conclusão de que uma diminuição nas compras pode ser resultado de uma diminuição nas pesquisas por produtos, que seria uma causa raiz desse risco.

É fundamental entender esse conceito de que cada ação percebida pode ser causada por eventos anteriores, pois na criação de *KRI's* espera-se que as causas raízes sejam mapeadas e monitoradas, como forma de antecipar possíveis problemas e resultados negativos antes que eles aconteçam. Como no exemplo apresentado, ao medir a quantidade de pesquisas por produtos, em tempo real, pode-se antecipar um potencial problema como a queda de vendas ou até mesmo uma grande procura por produtos que não possuem estoque o suficiente para atender a demanda.

Figura 2 – Linha do tempo de um risco



Fonte: NICHOLSON & BAKER (2013, p. 234)

A Figura 2 torna mais fácil o entendimento da cadeia de eventos que se sucedem desde a causa raiz até as consequências e impactos da materialização de um risco. É importante entender que os pontos à esquerda da percepção do risco, servem para a criação de *KRI*'s e à direita, para a mensuração das consequências e redução do impacto. É preferível ter a opção de agir antes que potenciais desvios aos resultados aconteçam.

A identificação das causas raízes pode ser feita com o auxílio de ferramentas da qualidade que auxiliam nessa detecção, como o diagrama de causa-efeito e a técnica “Os 5 Porquês”. Ambas técnicas podem ser encontradas em (SELEME, Robson, 2012). O mais importante nessa etapa é mapear o maior número de eventos anteriores àquele que resulta em um impacto aos resultados, possibilitando assim que medidas sejam tomadas com antecedência.

O grande desafio na criação de *KRI*'s está em identificar as causas raízes dos riscos mais críticos e estabelecer uma correlação entre dados históricos e atuais para gerar projeções assertivas que auxiliem os tomadores de decisão (NICHOLSON & BAKER, 2013). Como é impossível gerar dados sobre o futuro, deve-se utilizar a estatística e técnicas presentes na mineração de dados para lograr êxito nessa tarefa.

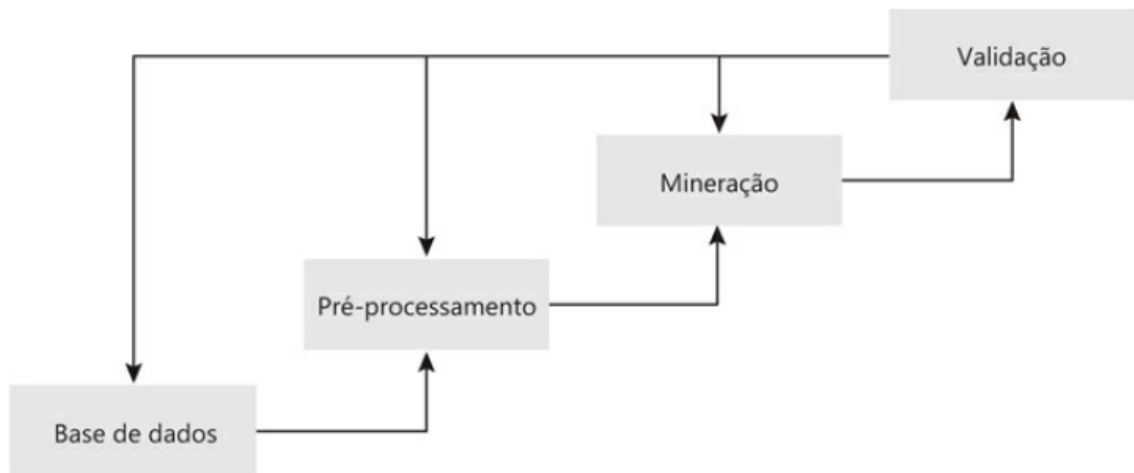
2.2 Mineração de Dados

A mineração, comumente conhecida, dispõe sobre a atividade de extrair material valioso diante de muitos materiais inúteis ou de pouca relevância, que após coletados passam por um processo de transformação até virarem produtos indispensáveis à vida humana contemporânea, como combustíveis, estradas, fertilizantes e equipamentos eletrônicos. Funcionando de forma parecida, diante dos avanços da tecnologia e imensidão de dados gerados a cada dia, a mineração de dados visa extrair informações valiosas e conhecimento, naquela que é conhecida como “a era dos dados”. Para Castro e Ferrari:

(...) esses avanços da tecnologia – tanto de hardware quanto de comunicação – têm produzido um problema de superabundância de dados, pois a capacidade de coletar e armazenar dados tem superado a habilidade de analisar e extrair conhecimento destes. Nesse contexto, é necessária a aplicação de técnicas e ferramentas que transformem, de maneira inteligente e automática, os dados disponíveis em informações úteis, que representem conhecimento para uma tomada de decisão estratégica nos negócios e até no dia a dia de cada um de nós. Nesse sentido, pesquisadores das mais variadas áreas têm se dedicado a estudar métodos para análise de dados. (2016, p. 41)

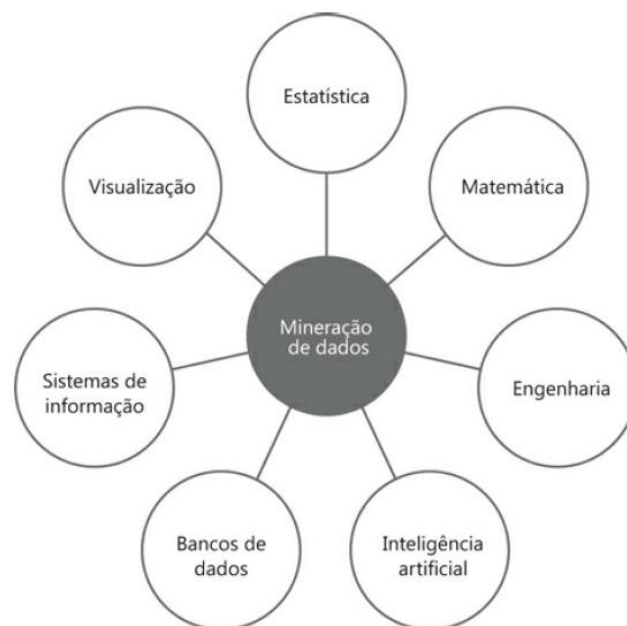
No contexto das organizações, a capacidade de lidar com uma grande quantidade de dados e obter informações relevantes e precisas, que contribuem com melhores decisões estratégicas e melhor distribuição de recursos, vem ganhando mais espaço. Profissionais como o analista de dados e o cientista de dados são muito disputados no mercado.

Para a construção de indicadores preditivos, o conhecimento de técnicas de mineração de dados e estatística é imprescindível, pois sem elas não é possível analisar dados existentes, realizar correlações e projetar o comportamento futuro, de forma assertiva e eficaz. “A mineração de dados é parte integrante de um processo mais amplo, conhecido como descoberta de conhecimento em bases de dados” (*knowledge discovery in databases*, ou KDD).” (CASTRO, Leandro & FERRARI, 2016). Conforme observa-se na Figura 3:

Figura 3 – Etapas do processo de KDD

Fonte: CASTRO & FERRARI (2016, p. 47)

O foco do presente trabalho é a etapa da Mineração de Dados, mas para o alcance de melhores resultados no processo de descoberta de conhecimento, deve existir uma etapa prévia de pré-processamento, para a remoção de dados inconsistentes e seleção de uma porção relevante para a fase de mineração, bem como uma etapa final de validação do conhecimento.

Figura 4 – Multidisciplinaridade da Mineração de Dados

Fonte: CASTRO & FERRARI (2016, p. 49)

A Figura 4 demonstra como a Mineração de Dados não é uma disciplina isolada, mas sim a reunião de diversas áreas de conhecimento, que juntas contribuem para melhores resultados. Para a construção dos indicadores preditivos, será dada uma maior atenção nos campos da estatística, matemática e inteligência artificial, no tocante ao aprendizado de máquina.

3 PROCESSO DE PREDIÇÃO DE DADOS

Prever o futuro é algo impossível, porém, através da junção de ferramentas específicas da Mineração de Dados, pode-se organizar os dados, reconhecer padrões, associar atributos e criar correlações que resultem em uma projeção de alta fidelidade que antecipe comportamentos de mercado, preferências dos clientes ou a taxa de inadimplência de uma região.

Figura 5 – Fluxo da criação de um modelo preditivo



Fonte: CASTRO & FERRARI (2016, p. 319)

A Figura 5 detalha o processo de aplicação de um modelo preditivo, desde a fase de pré-processamento de dados até a validação final.

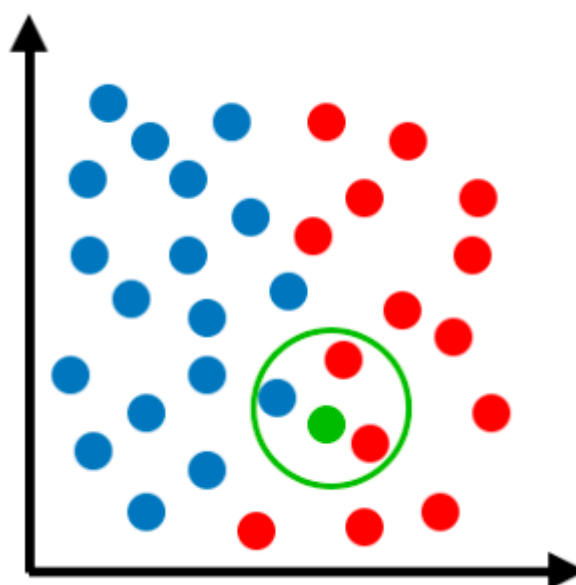
3.1 Ferramentas de predição de dados

As ferramentas de predição escolhidas para auxiliar na construção de indicadores preditivos foram o Classificador KNN, Árvores de Decisão e Regressão Linear.

3.1.1 Classificador KNN

O algoritmo Classificador KNN, do inglês *K Nearest Neighbor*, é um dos algoritmos preditivos mais importantes, sendo associado a projetos que envolvem aprendizado de máquina. Por ser fácil de aprender e simples de executar, pode ser aplicado na resolução de problemas em diversas áreas, principalmente quando há a necessidade de projetar a categoria de um dado novo. Tem a característica de algoritmo “preguiçoso”, pois não necessita de dados de treinamento prévios para a geração de um modelo, agilizando o processo inicial, em troca de mais atenção e cuidado na fase de testes e validação. (ABBOT, Dean, 2014).

Figura 6 – Funcionamento do Classificador KNN



Fonte: Do próprio autor do trabalho

Vê-se na Figura 6 que, diante de uma amostra de dados exibidos em forma de gráfico, o objetivo consiste em simular uma nova entrada, o ponto verde, e prever qual deve ser sua categoria. O algoritmo de classificação KNN necessita apenas que o usuário defina o valor da variável k , que indica a quantidade de pontos vizinhos

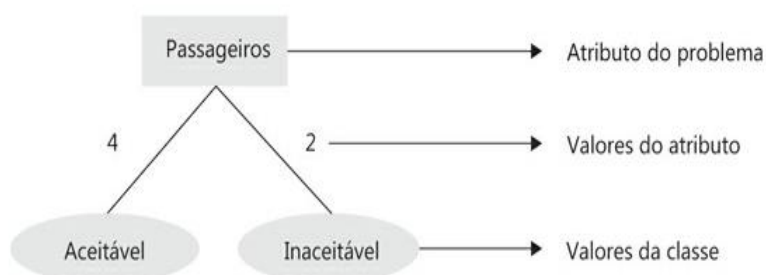
considerados para a análise. A partir de então, uma verificação é realizada em forma de contagem e a classe que possuir mais pontos naquela determinada região serve de referência. No exemplo supracitado, o valor de k é três, resultando em uma classe vermelha, pois na região avaliada há apenas um ponto azul e dois vermelhos.

Quanto maior a amostragem de dados, maior será a assertividade do algoritmo e menores serão as distorções e erros. Primariamente, o algoritmo é utilizado com dados qualitativos, categóricos nominais, mas através da utilização de técnicas estatísticas de regressão, deve realizar a predição de dados quantitativos, sem maiores problemas. (Referência)

3.1.2 Árvores de Decisão

De acordo com CASTRO & FERRARI, 2016, uma árvore de decisão é uma estrutura em forma de árvore na qual cada nó interno corresponde a um teste de um atributo, cada ramo representa um resultado do teste e os nós-folhas representam classes ou distribuições de classes. O nó mais elevado da árvore é conhecido como nó raiz, e cada caminho da raiz até um nó folha corresponde a uma regra de classificação, conforme a Figura 7.

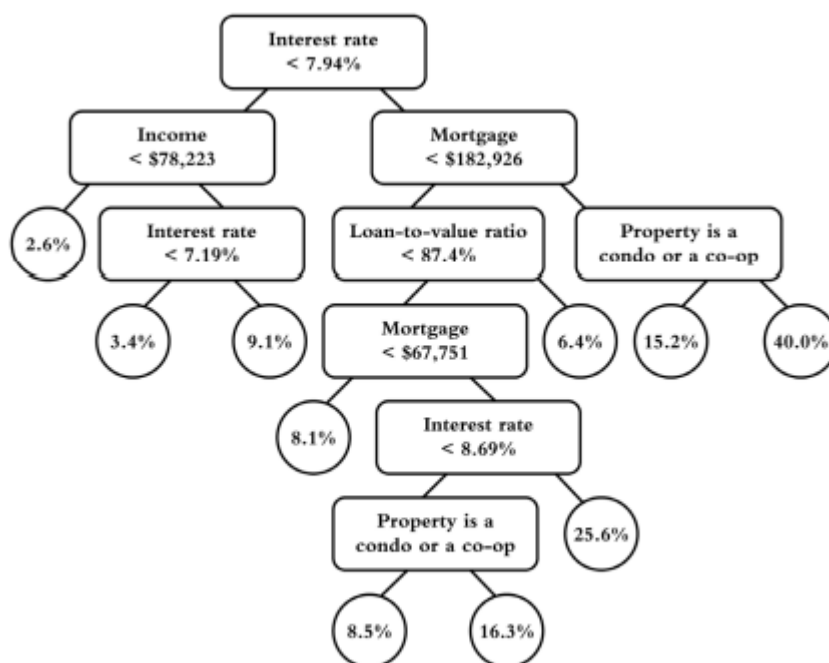
Figura 7 – Árvore de decisão simples para um carro



Fonte: CASTRO & FERRARI (2016, p. 343)

Segundo Siegel (2016), árvores de decisão são uma ótima forma de começar a trabalhar com mineração de dados e aprendizado de máquina, pois são mais fáceis aos olhos do que fórmulas matemáticas complexas. As árvores crescem de cima para baixo e encontram testes lógicos em nós, obedecendo respostas binárias, sim ou não.

Figura 8 – Árvore de decisão para taxa de juros

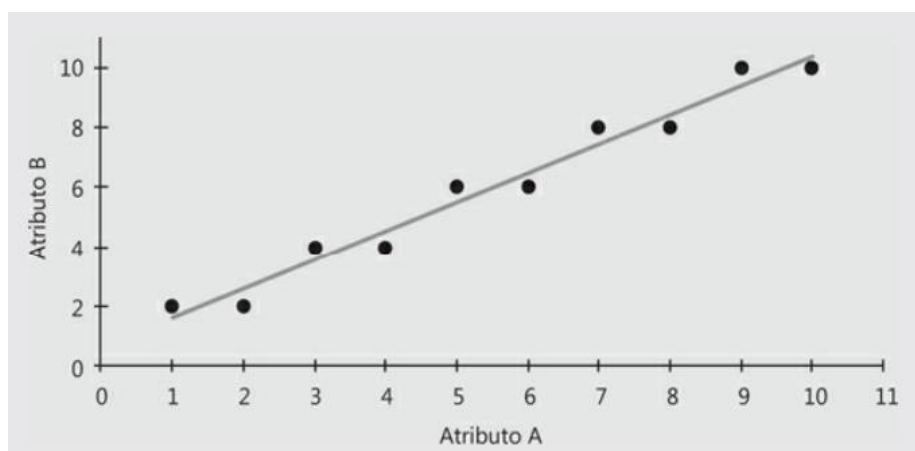


Fonte: Siegel (2014, p. 197)

Uma divisão é uma condição presente na árvore, onde os dados são divididos em dois ou mais subgrupos. Essas condições são similares aos testes condicionais presentes na fórmula SE, do Excel ou nas expressões condicionais presentes nas linguagens de programação. O dado é examinado, e caso a condição seja atendida, ele deve ir para o lado esquerdo, senão, para o direito. O fluxo continua até encontrar o último nó, que geralmente traz um percentual com a fatia da determinada população escolhida, conforme observa-se na Figura 8.

3.1.3 Regressão Linear

Segundo CASTRO E FERRARI, 2016, “Os modelos de regressão linear buscam relações entre duas variáveis por meio da determinação de uma equação de uma linha reta que melhor representa a relação entre as variáveis. Essa linha reta é denominada linha de regressão, ao passo que a equação é denominada equação de regressão”.

Figura 9 – Gráfico de Regressão Linear

Fonte: CASTRO & FERRARI (2016, p. 407)

A regressão linear procura pelos coeficientes da reta que minimizam a distância dos objetos à reta. Considere o conjunto de objetos apresentados na Figura 9, sendo cada objeto composto por dois atributos (A e B). A reta que mais se aproxima de todos os pontos apresenta a seguinte equação: $B = 0,9697 \times A + 0,6667$. Aplicando na equação os valores de A para cada objeto, obtém-se os valores estimados para B. A Tabela 6.4 apresenta os valores estimados (B') utilizando a equação e erro absoluto para cada objeto. O exemplo de regressão linear apresenta erro absoluto médio igual a 0,48.

Segundo Abbot (2014), a regressão linear consiste em traçar uma linha de tendência que melhor se adeque aos dados. É o algoritmo mais conhecido dentre os modelos de regressão e seu sucesso depende da escolha das variáveis. O modelo mais simples de seleção é chamado de “seleção para frente”, que funciona de uma forma parecida com as árvores de decisão, através da escolha de uma das trajetórias possíveis.

4 RESULTADOS E DISCUSSÃO

A finalidade do presente trabalho foi disseminar conhecimento sobre Mineração de Dados, sua multidisciplinaridade e importância no processo de descoberta de conhecimento em bases de dados e criação de indicadores preditivos.

Para tanto, foi utilizada a pesquisa bibliográfica e documental acerca dos principais métodos, algoritmos e ferramentas que permeiam o universo dessas disciplinas.

Segundo Falconi (2013), “somente aquilo que é medido é gerenciado. O que não é medido está à deriva”, portanto, é fundamental que as organizações estejam monitorando constantemente seus resultados. De acordo com Nicholson (2012), apenas monitorar o desempenho atual não é suficiente para agregar valor e sucesso, sendo necessário também um efetivo Gerenciamento de Riscos Corporativos, com um viés mais preditivo e com visão de futuro, perante os objetivos estratégicos de uma organização.

Ainda de acordo com Nicholson (2012), indicadores preditivos, também chamados de indicadores líderes, são ferramentas essenciais na etapa de monitoramento de resultados e vulnerabilidade perante as potenciais adversidades inerentes do negócio. O entendimento do funcionamento desses indicadores se fez necessário para que situações de risco e desvios aos objetivos sejam previstos o mais rápido possível, minimizando os impactos aos negócios.

Foi apresentado o conceito de Indicador Chave de Risco, funcionando de maneira parecida a um semáforo, tem como objetivo fornecer informações essenciais sobre a vulnerabilidade de uma organização perante eventos internos ou externos. O indicador possui três faixas de aceitação, divididas por dois gatilhos, que servem de alerta para a gestão e exigem uma tomada de decisão antecipada perante a distorções vinculadas às causas raízes dos riscos.

Tão importante quanto o modelo do Indicador Chave de Risco apresentado, estão os métodos e ferramentas utilizados para a obtenção de dados preditivos, através de análises e projeções envolvendo processamento de dados, algoritmos, estatística e aprendizado de máquina. Apenas com a união desses conhecimentos é possível construir modelos preditivos com altas taxas de acerto.

Para o presente estudo, as principais ferramentas escolhidas de acordo com maior afinidade com objeto de estudo foram o Classificador KNN, Árvores de Decisão e Regressão Linear. Segundo Castro & Ferrari (2016), a capacidade de criar um modelo com predição satisfatória é chamado de capacidade de generalização e, para alcançar tal sucesso, devem ser realizadas as etapas de treinamento e teste.

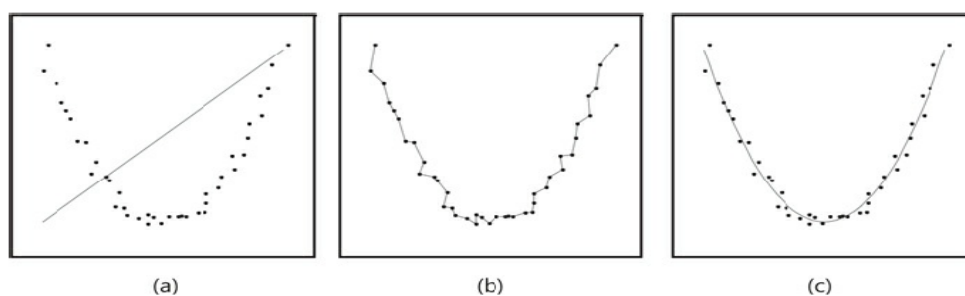
Na etapa de treinamento, os dados são utilizados como amostra e são definidas entradas e saídas desejadas para os registros selecionados. A ideia é que o modelo preditivo se torne cada vez mais assertivo, quanto mais amostras receber,

porém, segundo Castro & Ferrari (2016), os dados de treinamento podem sofrer distorções e desempenho deteriorado a partir de um determinado ponto. É válido lembrar que, algumas ferramentas, como o Classificador KNN, não precisam da fase de treinamento por possuir um funcionamento mais simples, porém, exigem uma etapa de validação e testes mais elaborada.

Ainda de acordo com Castro & Ferrari (2016), os dois principais erros na etapa de treinamento do modelo preditivo são o erro de representação e erro de generalização. O primeiro ocorre em decorrência do modelo não ser tão flexível a ponto de representar bem o conjunto de dados fornecidos, já o segundo ocorre quando o modelo é treinado em excesso e acaba absorvendo os ruídos da base de treinamento.

O ideal é encontrar um modelo preditivo equilibrado, com maior flexibilidade, sem absorver tanto os ruídos da base de treinamento, conforme ilustrado na Figura 10. A primeira parte, representada pela letra a, ilustra o erro de representação, a segunda, letra b, representa o erro de generalização e por fim, a letra c representa um modelo mais equilibrado.

Figura 10 – Erros comuns na etapa de treinamento de dados



Fonte: CASTRO & FERRARI (2016, p. 308)

Para Finlay (2014), a melhor prática e maneira de evitar distorções e erros comuns na etapa de treinamento, é manter uma terceira amostra separada, usada apenas para realizar a análise final de performance do modelo, evitando assim, possíveis vícios que as amostras utilizadas na fase de treinamento podem ocasionar. Outra boa prática é criar múltiplas amostras separadas com diferentes intervalos de tempo, testando assim a longevidade e confiabilidade de determinado modelo preditivo. É normal que modelos preditivos antigos percam sua efetividade com o

passar do tempo, portanto, é fundamental realizar análises periódicas sobre sua eficácia.

Na etapa de validação, os objetivos da validação cruzada são determinar quando interromper o treinamento dos modelos preditivos e estimar os erros citados anteriormente. Para cada rodada na validação cruzada é gerado um novo conjunto de treinamento e um de teste, quanto mais rodadas forem executadas, menor será a variabilidade nos resultados e maior será a probabilidade de acerto. (Steele; Chandler; Reddy, 2016).

Após a validação do modelo preditivo mais adequado para cada Indicador Chave de Risco, a próxima etapa é a inserção dos dados através da execução automatizada dos modelos. Na Figura 11, está disposto um modelo proposto de *KRI*, com três faixas, que representam o estado atual de cada risco e suas causas raízes e os limites de aceitação definidos. O visual deste modelo lembra um velocímetro, que mostra instantaneamente a velocidade do veículo para o motorista, e assim deve funcionar o *KRI*, sinalizando para os gestores qual é o nível de exposição ao risco em tempo real.



Figura 11 – Modelo de KRI



Fonte: Do próprio autor do trabalho

Na faixa verde, o risco está sob controle, em um patamar desejado. A faixa amarela representa um sinal de alerta para os responsáveis imediatos pela operação do processo envolvido. Já a faixa vermelha representa que o risco está prestes a ser materializado e exige que alguma ação seja tomada, sendo escalado imediatamente aos mais altos níveis organizacionais, como o conselho administrativo.

Figura 12 – Modelo de reporte de um Indicador Chave de Risco

| 1. Informação do KRI | 2. Risco | 3. Unidade de Medida | 4. Processo Relacionado | 5. Relógio de monitoramento | | | | | | | | | |
|--|------------------------|----------------------|----------------------------|---|---|--------|--------|--------|--------|--------|--------|--------|----------------|
| KRI 1 - Valor Total de Multas | ESTRATÉGICO | R\$ | Processo | KRI 1 - Valor Total de Multas  R\$ 15.000 Agir | | | | | | | | | |
| 6. Subprocesso | 7. Fórmula de Cálculo | | 8. Responsável | | | | | | | | | | |
| Subprocesso | Valor de Multas no Mês | | Responsável X | | | | | | | | | | |
| 9. Periodicidade | 10. Gatilhos | | 11. Comportamento Esperado | | | | | | | | | | |
| Mensal | R\$ 5.000 | R\$ 10.000 | R\$ 15.000 | |  Base Analítica | | | | | | | | |
| 13. Acompanhamento do indicador - 2022 | jan/22 | fev/22 | mar/22 | abr/22 | mai/22 | jun/22 | jul/22 | ago/22 | set/22 | out/22 | nov/22 | dez/22 | Acumulado 2022 |
| | R\$ 4.000 | R\$ 11.000 | R\$ 8.000 | R\$ 12.500 | R\$ 15.000 | | | | | | | | R\$ 13.000 |

Fonte: Do próprio autor do trabalho

Segundo Assi (2012), o monitoramento deve ser contínuo, periódico e seguir com o reporte aos gestores e à alta administração, exigindo que as devidas ações sejam tomadas de acordo com o nível de exposição a determinado risco. Na Figura 12, o indicador em questão encontra-se na faixa vermelha, exigindo que uma ação dos responsáveis seja tomada rapidamente, antes que a materialização do risco ocorra e traga consequências aos resultados.

5 CONSIDERAÇÕES FINAIS

O desenvolvimento deste trabalho possibilitou uma análise introdutória e melhor entendimento sobre a Mineração de Dados, sua multidisciplinaridade, processos, técnicas e importância na criação de modelos preditivos que servem de base para a implementação de indicadores preditivos, essenciais para o monitoramento contínuo do Gerenciamento de Riscos Corporativos.

Percebeu-se que indicadores de performance são importantes na busca por melhores resultados, mas que sem o auxílio de indicadores preditivos, associados aos riscos estratégicos de uma organização, não é possível gerenciar proativamente os recursos em direção aos resultados estratégicos. Para tanto, o Gerenciamento de Riscos Corporativos é a parte que identifica, avalia, monitora e reporta os riscos estratégicos através de indicadores preditivos, os KRI's.

Notou-se que a melhor maneira de criar indicadores líderes, que antecipam e projetam resultados é através de ferramentas da Mineração de Dados. Ao descrever e detalhar o processo de descoberta de conhecimento em base de dados, KDD, tornou-se possível entender a importância de cada etapa nesse processo de refinamento e definição de um modelo preditivo que seja mais adequado a cada um dos indicadores chave de riscos. Dessa forma, há menos distorções, erros, falhas e ruídos que diminuem consideravelmente a eficácia e sucesso desses modelos.

As ferramentas detalhadas no presente trabalho foram escolhidas por serem de fácil entendimento e aprendizado, necessitando de pouco conhecimento prévio, visando introduzir um público ainda não acostumado com os termos e recursos da Mineração de Dados. Cada uma das ferramentas tem papel fundamental na criação e implementação de modelos preditivos mais assertivos, garantindo uma maior previsibilidade nos resultados. Fez-se necessário apresentar erros e ruídos comuns resultantes do treinamento e execução dos modelos preditivos e as melhores formas de evitá-los, com o intuito de garantir uma maior eficácia e facilitar a validação de cada modelo.

Posterior à etapa de validação dos modelos preditivos, mostrou-se o momento de execução e alimentação dos indicadores de risco com os dados. Dessa forma, os indicadores compreendem informações valiosas acerca do nível atual de exposição às causas raízes que levam à materialização dos riscos da organização. Os indicadores devem ser monitorados constante e periodicamente e reportados aos responsáveis em tempo hábil, sempre que necessário, obedecendo a regra de funcionamento do KRI, ilustrada no presente trabalho.

Por fim, considerando o objetivo geral proposto, constatou-se que a Mineração de Dados é fundamental para a criação de modelos preditivos que auxiliam a implementação de indicadores preditivos, que garantem maior poder de decisão e antecipação e auxiliam o Gerenciamento de Riscos Corporativos no alcance dos objetivos estratégicos das organizações. Espera-se, com isso, contribuir para a construção de conhecimento sobre o tema, servindo de referencial para trabalhos futuros, bem como aplicações nas diversas áreas de conhecimento e trabalho.

REFERÊNCIAS

- ASSI, Marcos. **Gestão de Riscos com Controles Internos: ferramentas, certificações e métodos para garantir a eficiência dos negócios.** Local: São Paulo, Editora Saint Paul, 2012.
- FALCONI, Vicente. **Gerenciamento da rotina: do trabalho do dia a dia.** Local: Belo Horizonte, Editora Falconi, 2013.
- NICHOLSON, Francis & BAKER, Chris. **CRMA Certification in Risk Management Assurance.** Local: Altamonte Springs, Editora The IIA Research Foundation, 2013.
- CASTRO, Leandro & FERRARI, Daniel. **Introdução à Mineração de Dados: Conceitos básicos, algoritmos e aplicações.** Local: São Paulo, Editora Saraiva, 2016.
- ABBOT, Dean. **Applied Predictive Analytics: principles and techniques for the professional data analyst.** Local: Indianapolis, Editora Wiley, 2014.
- STEELE, Brian & CHANDLER, John & REDDY, Swarna. **Algorithms for Data Science.** Local: Cham, Suíça, Editora Springer, 2016.
- FINLAY, Steven. **Predictive Analytics, Data Mining and Big Data: myths, misconceptions and methods.** Local: Nova Iorque, Editora Palgrave Macmillan, 2014.
- SIEGEL, Eric. **Predictive Analytics: the power to predict who will click, buy, lie, or die.** Local: Nova Jersey, Editora Wiley, 2014.