



ANÁLISE DO TIPO DE PERFIL DO INVESTIDOR UTILIZANDO MINERAÇÃO DE DADOS

Investor profile type analysis using data mining

Guilherme Alan Pereira Rambo¹,

Patricia Mariotto Mozzaquatro Chicon²

Resumo: O tema mercado financeiro apresenta muito material disponível, tanto em bibliografias digitais quanto em livros. Apesar de elevada oferta de material disponível, a importância do tema está em reunir informações sobre o mercado financeiro, sobre os investimentos existentes no mercado e os perfis dos investidores. Muitos decidem investir sem dimensionar os riscos que estão expostos, e se esses riscos se encaixam no seu perfil. A mineração de dados por meio da descoberta de conhecimento pode auxiliar possíveis investidores na tomada de decisão. Neste sentido, a pesquisa busca apresentar um estudo sobre os tipos de perfis de investidores e aplicar a técnica de mineração de dados árvore de decisão a fim de identificar o perfil dos usuários (conservador, moderado ou agressivo), para assim, recomendar aplicações a serem realizadas. Dessa forma, um investidor pode tomar suas decisões com mais clareza, possuindo o risco de acordo com o que está disposto a correr, trazendo mais segurança e com isso, diminuir suas perdas. Para a geração desses perfis, foi utilizado o método da árvore de decisão por meio do algoritmo J48 a fim de classificar investidores conforme as características de cada perfil. A metodologia aplicada no estudo é quantitativa, foi aplicada a validação cruzada de dez partições a fim de verificar a acurácia do algoritmo de classificação. Quanto a realização dos testes, foram obtidos resultados positivos, foi constatada uma precisão de 85% por parte do algoritmo J48. Após constataremos a eficácia e precisão do algoritmo nos testes iniciais, foram realizadas previsões com usuários humanos, comprovando a assertividade do algoritmo..

Palavras-chave: Mineração de Dados. Investimentos. Perfil do investidor.

Abstract: The financial market theme has a lot of material available, both in digital bibliographies and in books. Despite the high offer of available material, the importance of the theme is to gather information about the financial market, about the investments existing in the market and the profiles of investors. Many decide to invest without dimensioning the risks they are exposed to, and whether these risks fit their profile. Data mining through knowledge discovery can help investors in decision making. In this sense, a research seeks to present a study on the types of investor profiles and apply a decision tree data mining technique in order to identify the profile of users (conservative, moderate or aggressive), in order to recommend applications to be accomplish. In this way, an investor can make his decisions more clearly, taking the risk according to what he is willing to take, bringing more security and thus reducing his losses. For the generation of profiles, the decision tree method was used through the J48 algorithm in order to

¹ Discente do curso de Ciência da Computação, da Universidade de Cruz Alta - Unicruz, Cruz Alta, Brasil.

E-mail: guilherme.rambo@hotmail.com

² Pesquisadora do Grupo Engetec, Docente da Universidade de Cruz Alta- Unicruz, Cruz Alta, Brasil.

E-mail: patriciamozzaquatro@gmail.com



classify investors according to the characteristics of each profile. The methodology applied in the study is quantitative, applied to cross-validation of ten partitions in order to verify the accuracy of the classification algorithm. As for the performance of the tests, positive results were obtained, with an accuracy of 85% by the J48 algorithm. After verifying the effectiveness and accuracy of the algorithm in the initial tests, predictions were made with human users, proving the assertiveness of the algorithm.

Keywords: Data Mining. Investments. Investor profile.

1 INTRODUÇÃO

A cada dia que passa, assuntos voltados ao mercado financeiro estão ficando mais populares, principalmente devido a alta da bolsa. O número de investidores pessoas físicas no Brasil cresceu 92% em 2020, chegando a um total de 3,2 milhões de contas (B3), representando um enorme crescimento da bolsa de valores. Antes da pandemia (Covid-19), a bolsa estava na margem dos quase 120 mil pontos, conforme fonte do ETF BOVA11, que tem como índice de referência o Ibovespa (INFO MONEY, 2020). O que ocorre, é que muitas pessoas estavam aderindo a bolsa de valores e investimentos, sem nenhum conhecimento ou assessoria, obtendo resultados negativos, vindo a perder seu dinheiro, e também não obtendo rendimentos satisfatórios de acordo com o risco que estava disposto a correr. Isto deve-se ao fato de não conhecer todos os investimentos disponíveis no mercado e quais seriam mais adequados para o seu perfil. Sendo assim, esta pesquisa busca identificar o perfil do investidor e apresentar os investimentos mais adequados.

Ao utilizar uma técnica computacional para avaliar o tipo de perfil do investidor, pode-se exibir quais os tipos de investimentos que se encaixam a ele, de acordo com o perfil apresentado. A principal vantagem está em auxiliar o investidor a obter um conhecimento maior sobre os tipos de investimentos que estão de acordo com seu perfil, visando uma melhor escolha e conforto em deixar seu dinheiro de forma segura e satisfatória de acordo com o ativo escolhido. Para que seja possível realizar todas as etapas do processo, será utilizada a mineração de dados.

1.1 Investimentos

De maneira ampla, o conceito de investimento é um desembolso em há que expectativa de ganho ou resultado futuro. Porém nem sempre somente capital pode ser investido, como também tempo, energia, estudo, atenção, dedicação e etc... Mas se tratando de investimento no mercado



financeiro, a ideia principal é aplicar dinheiro para que ele produza rendimentos futuros, devido aos juros compostos que trabalham a nosso favor nas aplicações financeira, fazendo com que o dinheiro se multiplique (BLOG MAGNETIS, 2020).

Um fato importante que devemos observar, é o fator “tempo”. Usando um conceito básico, na compra de um imóvel para alugar, geralmente o retorno é de médio/longo prazo, pois demorará um certo tempo para o imóvel ser pago com os alugueis, e somente a partir desse momento começará a dar lucro. Se tratando de investimentos em finanças, não é muito diferente, pois no mercado financeiro temos diferentes tipos de “tempo” (chamados de prazo) para os investimentos, sendo curto, médio e longo prazo. Não existe uma definição precisa sobre a duração do que é curto, médio ou longo prazo, mas alguns especialistas utilizam de 1 a 2 anos como curto, 3 a 9 anos como médio, e acima disso seria longo prazo. (VIDA E DINHEIRO, 2020).

Tratando-se a velocidade que um ativo pode ser convertido em dinheiro, chamamos esse fator de liquidez. Ou seja, uma liquidez grande é a facilidade de vender o ativo em qualquer momento, já baixa liquidez é uma dificuldade maior de vender, demorando mais tempo.

Outro fato importante a ser observado nos investimentos, é o fator “risco”. Em teoria, risco refere-se a algo não sair como o planejado, até porque não depende somente da vontade daqueles que estão diretamente ou indiretamente envolvidos, lembrando que o risco pode ser apenas minimizado, mas jamais eliminado (CAPELLETTO; CORRAR, 2007).

Outro fato importante é a rentabilidade. Em teoria, quanto maior o risco, maior será a rentabilidade, pois o usuário estará mais exposto ao risco, portanto deve ser melhor remunerado (caso seja). Também se têm o fator da liquidez, quanto maior for sua liquidez, menor será sua rentabilidade, pois terá mais facilidade em comercializar o ativo em questão. Portanto, a rentabilidade é uma mescla desses três principais fatores, além de outros.

Após apresentar a base do que são e como funcionam os investimentos, aborda-se o fundamento deste trabalho, que é analisar os tipos de perfis de investidores. Nesta pesquisa, iremos abordar três tipos de investidores, sendo eles conservador, moderado/arrojado e agressivo.

Conservador: Este tipo de perfil inclui investidores que não toleram riscos, ou toleram desde que seja um risco muito baixo, pois não querem arriscar perder seu patrimônio. Sendo assim, não estão muito preocupados com alta rentabilidade, pois os ativos que compõem sua carteira são de baixo risco, e em consequência, baixo retorno. (Caixa Econômica Federal, 2014). Geralmente visam o investimento para uma reserva de emergência, rendimento maior que deixar o dinheiro “parado” ou até mesmo alcançar metas menores. Para fugir da caderneta de poupança, e



também muitas vezes por estarem iniciando no mercado financeiro, os principais investimentos desse perfil estão em títulos de renda fixa, pois tem baixa variação e menor risco.

Moderado: A principal definição para este tipo de perfil, é aceitar correr riscos moderados. Ele está entre o conservador e o agressivo, estando disposto a correr um pouco mais de risco que o conservador, em busca duma maior rentabilidade, suportando perdas controladas. No caso de um moderado mais disposto a riscos, pode-se chamar de arrojado, que objetiva o longo prazo para incrementar seu patrimônio. Em geral, investidores desse perfil investem em fundos de investimentos, mesclando ou não a sua carteira com renda fixa. (SANTANDER, 2014).

Agressivo: Como o próprio nome já diz, é o perfil de maior agressividade se tratando de riscos. Este perfil está propenso a correr grandes riscos, em busca de uma maior rentabilidade (grandes lucros) e multiplicação de patrimônio. Alguns investidores desse perfil podem ser incoerentes, devido a não calcularem os verdadeiros riscos e ter noção dos prejuízos financeiros, comprometendo os lucros. Geralmente, a carteira deste perfil é formada por ações, além de realizarem operações no mercado futuro com o objetivo de ganhos e não de hedge (proteção). (XPI, 2014).

1.2 Mineração de Dados

A mineração de dados é uma tecnologia muito importante para centros de banco de dados, devido ao fato de que companhias podem gastar milhões de reais no trabalho de coleta de dados, e após essa coleta, nenhum retorno ou informação útil é encontrada em Larose, (2005). Sendo assim, a mineração de dados ajuda a obter retornos satisfatórios com os dados coletados, utilizando métodos que analisam os dados coletados e apresentam os resultados desejados.

Um desses modelos de transformação de dados em conhecimento, é a descoberta de conhecimento em base de dados (KDD), que surge como uma tentativa de solucionar o problema causado pela sobrecarga de dados. Na maioria dos casos, devido ao enorme volume de dados, o processo manual torna-se impraticável, e por isso é utilizado este processo.

Pode-se dividir o KDD em três grandes etapas operacionais: etapa de pré-processamento, etapa de Data Mining (Mineração de Dados) e etapa de pós-processamento. (GOLDSCHMIDT; PASSOS, 2005),

A etapa do pré-processamento de dados compreende as funções que se relacionam a captação, à organização e ao tratamento de dados, cujo objetivo é preparar os dados para os algoritmos da etapa seguinte (data mining). A etapa de mineração de dados realiza a busca efetiva



de conhecimentos úteis no contexto proposto para a aplicação do KDD, sendo a principal etapa do processo de KDD. Compreende a aplicação de algoritmos sobre os dados procurando abstrair o tão esperado conhecimento. Também pode ocorrer por meio de diferentes métodos e cada um deles requer diferentes necessidades de pré-processamento. A etapa de pós processamento abrange o tratamento do conhecimento obtido na etapa anterior (Data Mining), envolvendo a visualização, análise e interpretação do modelo gerado. Tem como objetivo a facilidade de interpretação e da avaliação, por parte do homem, em relação á utilidade do conhecimento descoberto (GOLDSCHMIDT; PASSOS, 2005),

2.2.1 Tarefa de classificação

Atualmente, pode-se classificar a mineração de dados pela sua capacidade em realizar determinadas tarefas, como por exemplo: Descrição, Classificação, Estimação ou Regressão, Predição, Agrupamento e Adição. Nesta pesquisa, usaremos a tarefa de Classificação como forma de atribuir o perfil do cliente ao grupo relacionado (tipo de investidor).

Em resumo, a tarefa de Classificação objetiva direcionar para qual grupo um registro de dados deve pertencer, sendo uma das tarefas mais comuns da mineração de dados. Para isso, o modelo analisa os novos dados fornecidos (registro), com cada registro já contendo sua classe, fazendo com que ocorra o “aprendizado supervisionado”, que é nada mais que “aprender” a classificar um novo registro. (CAMILO, 2009).

A classificação tem inúmeras aplicações, incluindo detecção de fraudes, marketing, previsão de desempenho, fabricação, diagnóstico médico e sistemas de gestão (HAN et al., 2011).

A classificação pode ou não ser supervisionada, sendo utilizada principalmente para classificar registros em suas respectivas categorias. A classificação apresenta alguns métodos, como por exemplo: Árvore de Decisão, Classificação Bayesiana, Classificação Baseada em Regras, Redes Neurais., Algoritmo Genético, Classificação por regras de associação, SVM.

A árvore de decisão é um método de classificação que apresenta como principal objeto um fluxograma em forma de árvore, onde cada nó (não folha) indica um teste feito sobre um determinado valor (por exemplo, valor < 40). As ligações existentes entre os nós significam os possíveis valores do teste do nó superior, e as folhas indicam a classe (categoria) na qual pertence o registro (GAMA, 2000), sendo uma das técnicas de aprendizado de máquina onde um problema complexo é decomposto em subproblemas mais simples, e assim, a mesma estratégia é aplicada em cada subproblema (GAMA, 2000).



Após montada a árvore de decisão, para classificarmos um novo registro basta seguirmos o fluxo na árvore, começando do nó raiz até chegar numa folha. Por possuir esta estrutura, que as árvores de decisão podem ser definidas como uma regra de classificação. Além disso, o sucesso do método de árvore de decisão é devido a ser algo extremamente simples, pois não necessita de parâmetros de configuração e também apresenta um bom grau de assertividade. (CAMILO, 2009).

São aplicados métodos específicos (algoritmos de aprendizagem de máquina) para extração de padrões a partir dos dados ou até mesmo revelar o comportamento de um banco de dados. A seguir é descrito o algoritmo J48 mais citado, o qual implementa a técnica de classificação (GOLDSCHMIDT et al., 2015).

O J48 é um algoritmo de código aberto, implementado pelo software WEKA (Witten et al.; 2016), que também utiliza a estratégia “dividir para conquistar”, baseando-se no conceito de razão de ganho de informação que idêntica por meio da redução de entropia o quanto informativo um atributo é, para então selecionar a separação ótima, ou seja, o quanto espera-se que a entropia se reduza caso um determinado nó seja escolhido para fazer a partição dos dados (LAROSE, 2005).

Além de trabalhar com atributos qualitativos, contínuos e discretos, como citado anteriormente, o algoritmo J48 também é capaz de trabalhar com valores categóricos e ausentes. O tratamento de atributos contínuos envolve a consideração de todos os valores presentes no conjunto de treinamento, fazendo com que sejam ordenados de forma crescente considerando todos os valores presentes nos dados de treinamento e, após esta ordenação, seja selecionado o valor que favorecerá a redução da entropia (CAMARGO et al., 2016); (RAMYA et.al, 2015).

2 PROCEDIMENTOS METODOLÓGICOS

A seguinte pesquisa tem natureza aplicada, ou seja, objetiva gerar conhecimentos para aplicação prática dirigida à solução de problemas específicos. Envolve verdades e interesses locais. GIL (2002). Quanto aos Objetivos é uma Pesquisa Exploratória, já quanto aos Procedimentos é uma Pesquisa Experimental. A pesquisa foi dividida nas 3 seguintes etapas: Estudo Teórico, desenvolvimento e validação.

Etapas 1 – Estudo teórico

Estudar sobre perfis de investidores;

Relacionar os perfis com grupos de investimentos;



Pesquisar sobre técnicas e métodos de mineração de dados.

Etapa 2 – Desenvolvimento

Criação da base de dados;

Criação do arquivo arff;

Aplicação do algoritmo de mineração de dados;

Geração dos perfis;

Testes iniciais

Etapa 3 – Validação

criação da interface gráfica e integração com a ferramenta Weka;

testar os modelos de perfis gerados.

3 RESULTADOS E DISCUSSÃO

Para o desenvolvimento do trabalho foram utilizados os seguintes recursos de hardware e software. Como recursos de hardware, foi utilizado um Notebook com processador i5, 8 gb ram e Windows 10, sendo recurso próprio do autor. Como recursos de software, foi utilizada a ferramenta Weka para geração de perfis, XAMPP, MySQL e Linguagem PHP para a geração da base de dados e por fim, Star UML para geração dos diagramas.

Inicialmente, foi realizado um treinamento através da geração de uma base de dados fictícia em PHP, contendo dados randômicos, ou seja, possíveis respostas de 1025 usuários, a fim de treinar o algoritmo e verificar seus resultados. Para esse treinamento, foram utilizados 6 atributos e 1 classificador.

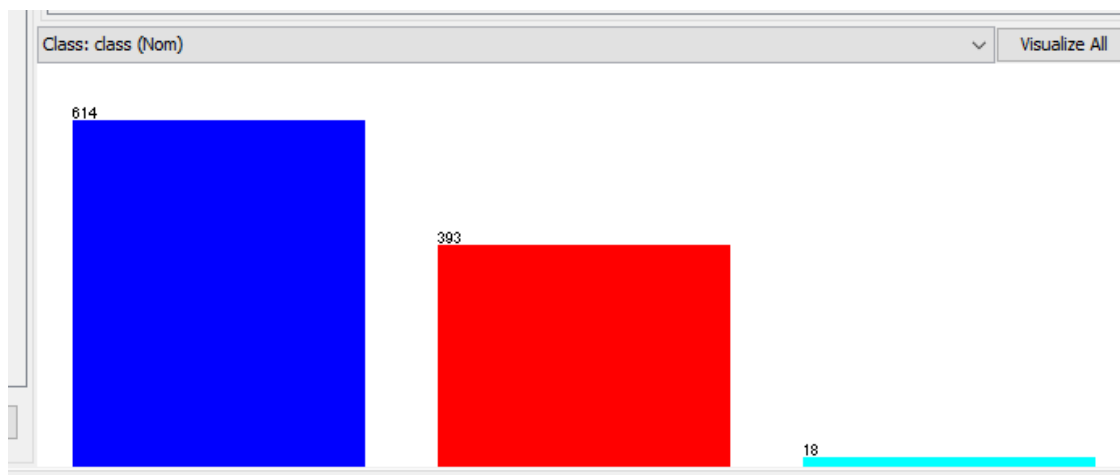
Para gerar a base, foi criado um arquivo em PHP com 6 perguntas (atributos), com respostas randômicas. Cada resposta possui um valor de atributo, sendo que quando mais próximo do zero, mais conservador o perfil, e quanto maior o valor do atributo, mais agressivo o perfil possivelmente será. Sendo assim, foram utilizados valores fixos para representar se o perfil do usuário será conservador, moderado ou agressivo.

Para realizar as predições foi utilizada a ferramenta Weka. Dentro do Weka, foi utilizado a classificação baseada em árvore de decisão, e após foi escolhido o algoritmo j48 para geração da



base. A base de treinamento tem um total de 1025 registros, sendo que 814 apresentaram o perfil agressivo, 392 apresentaram perfil moderado e apenas 18 com perfil conservador, conforme a Figura 1.

Figura 1 – Base de treinamento



Fonte: Autor

Foram realizadas algumas predições sem inserir dados em alguns atributos, a fim de realizar testes com o algoritmo. Para testar a assertividade, foram realizados alguns testes, modificando respostas dos atributos de usuários que já foram identificados os perfis. Para esse teste, foram inseridos pontos de interrogação “?” no lugar das respostas, ou seja, além de não identificar o perfil, também não foram identificados alguns campos dos atributos, a fim de verificar se mesmo com alguns campos modificados, o algoritmo irá retornar corretamente o perfil de investidor do usuário (Figura 2).

Figura 2- Predições realizadas

```

1 @relation investimentos_CERTO
2
3 @ATTRIBUTE QUESTAO1 {0,1}
4 @ATTRIBUTE QUESTAO2 {0,1,2}
5 @ATTRIBUTE QUESTAO3 {0,1,2,3,4}
6 @ATTRIBUTE QUESTAO4 {0,1,2}
7 @ATTRIBUTE QUESTAO5 {0,1,2,3}
8 @ATTRIBUTE QUESTAO6 {0,1,2}
9
10 @ATTRIBUTE class{AGRESSIVO, MODERADO, CONSERVADOR}
11
12
13
14 @data
15 1,2,4,2,?,?,?
16 1,?,3,?,3,1,?
17 ?,?,?,?,2,2,?
18 ?,?,3,?,3,2,?
19 1,?,3,2,?,1,?
20 ?,2,1,?,2,?,?
21 1,2,1,1,1,1,?
22 1,?,2,1,1,?,?
23 1,2,4,1,3,?,?
24 1,2,1,2,2,2,2
    
```

Fonte: Autor



Após foram realizadas predições com usuários reais, através da criação de um formulário, onde as respostas foram inseridas de forma manual na ferramenta Weka. O questionário foi elaborado no google forms, onde cada usuário acessou através de um link.

Foi aplicada a validação cruzada de dez partições em que os dados são divididos em dez partições aleatórias sendo retiradas nove dessas partições para serem utilizados no conjunto de treinamento e uma partição para o conjunto de testes. Dessa primeira iteração é obtida a primeira precisão do modelo. Após são realizadas, para o algoritmo de classificação aplicado, mais nove iterações, percorrendo todas as possibilidades de escolha, resultando em mais nove valores de precisão (ROMERO et al., 2013) (Figura 3).

Figura 3- Validação do algoritmo aplicado - J48

Correctly Classified Instances	875	85.3659 %
Incorrectly Classified Instances	150	14.6341 %
Kappa statistic	0.6998	
Mean absolute error	0.1342	
Root mean squared error	0.295	
Relative absolute error	40.7078 %	
Root relative squared error	72.7145 %	
Total Number of Instances	1025	

Fonte: Autor

Após a realização de todos os testes, foi identificado que o algoritmo apresentou um alto grau certeza ou instâncias corretas de pelo menos 85% dos resultados. O teste foi realizado com 8 usuários, pelo menos 6 deles estão com o perfil de investidor correto.

Sendo assim, o algoritmo irá entregar pelo menos 85% dos resultados com precisão, podendo ser aumentado conforme a quantidade de instâncias e testes realizados. A fim de explicar os resultados da validação, alguns conceitos sobre as métricas são descritos. Kappa: É o valor calculado dos resultados encontrados nos registros e relatado como um decimal (0 a 1). Quanto menor o valor de kappa menor a confiança da observação, o valor 1 implica a correlação perfeita. Mean absolute error: É a média de diferença entre os valores atuais e os preditos em todos os casos, ou seja, é a média do erro da predição. Root mean squared error: É utilizado para medir o sucesso de uma predição numérica. Apresenta o valor de erro entre os valores atuais e os valores preditos. Relative absolute error: É o erro total absoluto. Em todas as mensurações de erro, valores mais baixos significam maior precisão do modelo, com valor próximo de zero temos o modelo estatisticamente perfeito. (ROMERO et al., 2013).



Na base de treinamento e validação cruzada, foram classificadas 875 instâncias corretamente com um percentual de 85.36%. O Kappa obteve um valor aproximado de 0,7, o que é considerado bom (próximo de 1). O Mean Absolute Error (0.13) e o Roat Mean Squared Error (0.29) apresentaram valores consideráveis, pois são baixos. Já o Relative Absolute Error, apresentou um valor mais elevado de 40% aproximadamente, tendo em vista que o melhor valor deveria ser o mais próximo de 0 possível. Concluindo, analisando todas as métricas, o algoritmo apresentou uma boa acurácia na validação.

4 CONSIDERAÇÕES FINAIS

Esta pesquisa é parte integrante de um trabalho de conclusão de curso em andamento. O principal objetivo é informar qual o perfil de investidor que o usuário possui, e de acordo com os processos de mineração de dados que foram realizados, como por exemplo o treinamento da base que será usada e as predições geradas, com algumas respostas aleatórias a fim de testar a veracidade do algoritmo. Os resultados foram apresentados de forma positiva, com uma taxa de assertividade acima de 85%, o que torna o algoritmo válido para identificar qual perfil de investidor o usuário possui.

Até o momento, para obter os dados de usuários reais, foi utilizado um formulário no google forms, as respostas foram inseridas manualmente no algoritmo pelo autor, a fim de testar as predições se estavam corretas. E os resultados foram positivos.

Posteriormente, ainda será criado um formulário mais elaborado em PHP, onde não será preciso inserir os dados de forma manual, pois esse formulário será integrado a ferramenta Weka, que está sendo utilizada neste trabalho.

REFERÊNCIAS

BLOG MAGNETIS. Disponível em:< <https://blog.magnetis.com.br/juros-compostos/>>. Acesso em Dez de 2020.

CAIXA ECONÔMICA FEDERAL. Perfil do investidor. Disponível em:< http://www11.caixa.gov.br/portal/public/investidor/investidor/invista/perfil_do_investidor>. Acesso em Dez de 2020.

CAMARGO, A., SILVA, R., AMARAL, É., HEINEN, M. and PEREIRA, F. - Mineração de dados eleitorais: descoberta de padrões de candidatos a vereador na região da campanha do Rio Grande do Sul, Revista Brasileira de Computação Aplicada 8(1): 64–73. 2016.



CAMILO, Cássio Oliveira; SILVA, João Carlos da - Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas - Relatório Técnico, 2009.

CAPELLETTO, Lucio Rodrigues; CORRAR, Luiz João. Índices de risco sistêmico para o setor bancário. Revista Contabilidade & Finanças, v. 19, n. 47, p. 6-18, 2008.

FAYYAD, U; PIATETSKY-SHAPIRO, G; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence, 1996.

GAMA, J. Árvores de Decisão, 2000. Disponível em:
<<http://www.liacc.up.pt/~jgama/Mestrado/ECD1/Arvores.html>>. Acesso em: jun. 2020.

GIL, Antônio Carlos, 1946- Como elaborar projetos de pesquisa/Antônio Carlos Gil. -4. ed. -São Paulo: Atlas, 2002.

GOLDSCHMIDT, R., BEZERRA, E. & PASSOS, E. (2015), Data mining: conceitos, técnicas, algoritmos, orientações e aplicações, 2015, Rio de Janeiro-RJ: Elsevier pp. 56–60.

GOLDSCHMIDT, R.R.; PASSOS, E. Data Mining: Um Guia Prático. Rio de Janeiro: Campus, 2005.

HAN, J., KAMBER, M., & PEI, J. Data Mining: Concepts and Techniques.1485(3rd ed.). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. (2011).

INFO MONEY – Disponível em < www.infomoney.com.br> Acesso em nov. 2020.

LAROSE, D. T. Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley and Sons, Inc, 2005.

RAMYA, M., LOKESH, V., MANJUNATH, T. and HEGADI, R. S. (2015). A predictive model construction for mulberry crop productivity, Procedia Computer Science 45: 156–165.

ROMERO, Cristóbal et al. Predicting students' final performance from participation in on-line discussion forums. Computers & Education, v. 68, p. 458-472, 2013.

SANTANDER. Perfil do investidor. Disponível em:
http://www.santander.com.br/portal/gsb/gcm/package/asset_management/suitability_49179.zip/home.htm>. Acesso dez. 2020

SPÓSITO, M. A., de Castro, T. H. & de Castro Jr, A. N. (2008), Estação de Percepção: uma abordagem para o monitoramento em ambientes virtuais de aprendizagem, in 'Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), 2008. Vol. 1, pp. 288–298.

