

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

THIAGO PROENÇA MEIRELLES CORREA RAPHAEL

**ANÁLISE DA PREVISÃO DE RETORNO DOS PRINCIPAIS FUNDOS DE RENDA
FIXA BRASILEIRA, ATRAVÉS DE ALGORÍTMOS DE APRENDIZAGEM DE
MÁQUINA.**

LONDRINA

2021

THIAGO PROENÇA MEIRELLES CORREA RAPHAEL

**ANÁLISE DA PREVISÃO DE RETORNO DOS PRINCIPAIS FUNDOS DE RENDA
FIXA BRASILEIRA, ATRAVÉS DE ALGORÍTMOS DE APRENDIZAGEM DE
MÁQUINA.**

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Bacharel em Engenharia de Produção, do Departamento de Engenharia de Produção, da Universidade Tecnológica Federal do Paraná.

Orientador: Profa. Dr. Rogério Tondato

LONDRINA

2021

THIAGO PROENÇA MEIRELLES CORREA RAPHAEL

**ANÁLISE DA PREVISÃO DE RETORNO DOS PRINCIPAIS FUNDOS DE RENDA
FIXA BRASILEIRA, ATRAVÉS DE ALGORÍTMOS DE APRENDIZAGEM DE
MÁQUINA.**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do título
de Bacharel em Engenharia de Produção da
Universidade Tecnológica Federal do Paraná
(UTFPR).

Data de aprovação: 19/agosto/2021

José Ângelo Ferreira
Doutor
Universidade Tecnológica Federal do Paraná

Silvana Rodrigues Quintilhano
Doutora
Universidade Tecnológica Federal do Paraná

Rogério Tondato
Doutor
Universidade Tecnológica Federal do Paraná

LONDRINA

2021

RESUMO

O mercado financeiro possui um número cada vez mais elevado de fundos de investimento, tornando-se mais acessível ao público em geral, com suas infinitas possibilidades e baixos valores de entrada. Ao mesmo tempo com tantas oportunidades em aberto, ele possui as peculiaridades individuais de cada uma delas, complexidade nos nomes, rentabilidades, variações e taxas. Dificultando muito para a população compreender o mesmo e entender como as movimentações podem ocorrer com base na economia atual. Neste contexto, o presente trabalho teve como objetivo analisar dois fundos de investimento e verificar a previsão de retorno através de algoritmos de aprendizagem de máquina, relacionados a três grandes indicadores nacionais: Câmbio, CDI e IPCA. Para tal foi utilizado algoritmo Naive Bayes na ferramenta Weka (*Waikato Environment for Knowledge Analysis*), Regressão Linear via Excel e a Planilha de Previsão via Excel. Tomando como base os resultados nos anos de 2018, 2019 e 2020. Os resultados obtidos foram que não há relação entre a variação dos fundos e os preditores determinados, mostrando que a previsão por séries temporais ainda é a mais eficaz, mesmo que distante de um bom resultado.

Palavras-chave: Fundo de Investimento. Renda Fixa. Câmbio. CDI. IPCA. Naive Bayes, Regressão Linear, Planilha de Previsão.

ABSTRACT

The financial market has an increasing number of investment funds, becoming more accessible to the general public, with its infinite possibilities and low entry values. At the same time with so many open opportunities, it has the individual peculiarities of each one of them, complexity in names, profitability, variations and rates. Making it very difficult for the population to understand the same and understand how movements can occur based on the current economy. In this context, the present work aimed to analyze two investment funds and verify the return forecast through machine learning algorithms, related to three major national indicators: Exchange, CDI and IPCA. For this purpose, the Naive Bayes algorithm was used in the Weka tool (Waikato Environment for Knowledge Analysis), Linear Regression via Excel and the Forecast Worksheet via Excel. Based on the results for the years 2018, 2019 and 2020 . The results obtained were that there is no relationship between the variation of funds and the determined predictors, showing that the forecast by time series is still the most effective, even if far from a good result.

Keywords: Investment fund. Fixed Income. Exchange. CDI. IPCA. Naive Bayes. Linear Regression. Forecast Worksheet.

LISTA DE FIGURAS

Figura 1 - Participação de mercado no estoque de crédito - Pessoas físicas	33
Figura 2 - Condições de Investimento – CAIXA ECONÔMICA FEDERAL.....	34
Figura 3 - Condições de Investimento – BANCO DO BRASIL	35
Figura 4 - Resultados Método de Naive Bayes – Fundo Caixa.....	36
Figura 5 - Resultado Método Regressão Linear - Fundo Caixa vs indicadores	37
Figura 6 - Resultado Método Regressão Linear - Fundo Caixa vs CDI.....	38
Figura 8 - Dados reais e previstos pela Planilha de Previsão, aplicado a base de 36 meses – Fundo Caixa	39
Figura 9 – Dados Previstos Através da Planilha de Previsão – Fundo Caixa	39
Figura 10 - Resultados Método de Naive Bayes – Fundo BB	40
Figura 11 - Resultado Método Regressão Linear - Fundo BB vs indicadores	41
Figura 12 - Resultado Método Regressão Linear – Fundo BB vs CDI	42
Figura 13 - Dados reais e previstos pela Planilha de Previsão, aplicado a base de 36 meses – Fundo BB.....	43
Figura 14 - Dados Previstos Através da Planilha de Previsão – Fundo BB	43

LISTA DE TABELAS

Tabela 1 - Ranking de Administração de Fundos de Investimento	12
---	----

SUMÁRIO

1. INTRODUÇÃO	7
1.1 OBJETIVOS.....	8
1.2 JUSTIFICATIVA	8
1.3 MÉTODOS E TÉCNICAS DE PESQUISA.....	9
2. REFERENCIAL TEÓRICO.....	11
2.1 FUNDOS DE INVESTIMENTOS	11
2.2 ÍNDICE NACIONAL DE PREÇOS AO CONSUMIDOR AMPLO.....	16
2.3 TAXA DE CÂMBIO	18
2.4 TAXA DE JUROS (CDI).....	20
3. MODELOS DE ANÁLISE DE DADOS PARA SÉRIES ECONÔMICAS	22
3.1 CLASSIFICAÇÃO	22
3.1.1 <i>Aprendizagem Bayesiana</i>	22
3.1.2 <i>Por arvores de decisão</i>	23
3.1.3 <i>Por regras</i>	23
3.1.4 <i>Instâncias</i>	24
3.1.5 <i>Regressão logística</i>	25
3.1.6 <i>Máquinas de vetores de suporte (MVS)</i>	25
3.1.7 <i>Redes neurais artificiais</i>	26
3.2 REGRESSÃO LINEAR	27
3.2.1 <i>Não linear</i>	27
3.3 AGRUPAMENTO.....	28
3.3.1 <i>K-means</i>	28
3.3.2 <i>Hierárquico</i>	28
3.4 ASSOCIAÇÃO.....	29
3.4.1 <i>A priori</i> 29	
3.4.2 <i>ECLAT</i> 30	
4. DESENVOLVIMENTO DO TRABALHO	32
4.1 WEKA COMO SUPORTE A ANÁLISE DE DADOS.....	32
4.2 DEFINIÇÃO DOS FUNDOS A SEREM INVESTIGADOS	32
4.3 DEFINIÇÃO DOS INDICADORES A SEREM COMPARADOS.	35

5. RESULTADOS E DISCUSSÕES.....	36
5.1 ANÁLISE DE DADOS PARA O FUNDO CAIXA.....	36
5.2 ANÁLISE DE DADOS PARA O FUNDO BB.....	40
6. DISCUSSÕES.....	44
7. CONCLUSÃO.....	45
8. REFERÊNCIAS.....	46

1. INTRODUÇÃO

Os fundos de renda fixa são considerados uns dos mais seguros do mercado, porém se faz necessário analisar, como os mesmos variam conforme a economia está situada no momento, para um maior entendimento sobre como investir e os níveis de retorno que se tem com base na economia atual. Uma grande dificuldade para o público que não está inserido no mercado é a compreensão das diversas variações tanto positivas quanto negativas das aplicações econômicas financeiras de renda fixa ao longo do tempo, como as notícias econômicas afetam os investimentos pessoais, como as tomadas de decisões do governo impactam seus rendimentos. Desta forma, procura-se entender através de algoritmos de aprendizado de máquina se existe uma relação da previsão de retorno entre principais fundos de renda fixa brasileira, e três conhecidos indicadores econômicos nacionais.

Entender como e se as mínimas variações dos fundos têm relação ou não com alguns índices importantes do mercado, trazendo a tona uma realidade de que se prestarmos atenção no dia-a-dia das tomadas de decisões do governo, de grandes conglomerados econômicos, poderemos perceber variações na nossa carteira de investimentos. Com um conhecimento mais amplo nesse sentido, possibilitará ao pequeno investidor diversificar seu capital e se precaver para grandes impactos que podem vir a ocorrer no mercado, como uma frase dita por Robert Brokamp “Diversificação reduz os riscos, aumenta a previsibilidade e impulsiona os retornos.”.

Segundo Cardozo (2019) no Brasil, em pesquisa realizada pela Associação Brasileira das Entidades dos Mercados Financeiros e de Capitais Anbima (2019), a qual realizou entrevistas com pouco mais de 3 mil pessoas ao longo do Brasil, 58% dos brasileiros declaram que não possuem nenhum investimento. De acordo com a pesquisa as principais razões que levam o brasileiro a não investir estão relacionados baixo salário/sem condições (63%), desempregado(a)/não tem emprego fixo (10%) ou gastos inesperados (7%), correspondendo a 80% do número total de entrevistado

1.1 Objetivos

O objetivo geral deste trabalho é verificar a possibilidade de prever o retorno dos principais fundos de renda fixa através de algoritmos de aprendizagem de máquina.

Como objetivos específicos, tem-se:

- Apresentar um referencial teórico sobre o tema;
- Identificar e levantar os dados obtidos de variações dos fundos e dos indicadores a serem utilizados no trabalho;
- Aplicar modelos de análise de dados para verificar a existência de relação entre indicadores financeiros e aplicações financeiras;
- Analisar os resultados obtidos e divulgar as possíveis relações.

1.2 Justificativa

Observando o mercado atual de investimentos no Brasil, pode-se perceber que o número de brasileiros que tem algum tipo de investimento ainda é muito baixo. Segundo dados da ANBIMA (2019), em 2018 o percentual de brasileiros que possuíam algum valor aplicado em produtos de investimento era de 42%, mesmo percentual do ano anterior. Entretanto, o levantamento também traz que dos 33% dos brasileiros que economizaram em 2018, 48% aplicaram em produtos financeiros – em 2017 o percentual era de 42%, um leve aumento. Analisando esses números, podemos ver que ainda temos muito o que evoluir dentro desse mercado e um dos grandes fatores para esses números não subirem, é a falta de conhecimento e entendimento de como as relações do governo e economia afetam sua carteira de investimento.

Por meio desse trabalho, procura-se verificar se há possíveis relações entre índices econômicos facilmente reconhecidos pela população brasileira, com a rentabilidade dos fundos de renda fixa, podendo assim clarear e auxiliar as pessoas a entrarem neste mercado, que vem crescendo ainda há passos pequenos.

1.3 Métodos e Técnicas de Pesquisa

Quanto à natureza da pesquisa é quantitativa, pois serão estudados os indicadores econômicos e as aplicações financeiras.

Segundo (SOUSA, 2006 p.2) desenhos de pesquisa quantitativos geralmente refletem uma filosofia determinista que está baseada no paradigma ou escola de pensamento pós-positivista. Pós-positivistas examinam causa, e como diferentes causas interagem e/ou influenciam resultados. O paradigma pós-positivista adota a filosofia de que a realidade pode ser descoberta; entretanto, apenas imperfeitamente e num senso probabilístico. A abordagem é tipicamente dedutiva onde a maioria das ideias ou conceitos é reduzida a variáveis e as relações entre elas são testadas. O conhecimento que resulta é baseado em observação, medição e interpretação cuidadosas da realidade objetiva.

A pesquisa é explicativa, pois segundo (DAMY, 2002, p. 3). “como preocupação central identificar os fatores que determinam ou que contribuem para a ocorrência dos fenômenos. Pode-se dizer que o conhecimento científico está assentado nos resultados oferecidos pelos estudos explicativos”

Quanto ao método de pesquisa constitui-se na modelagem. Para Burak (1992, p.62),

A Modelagem Matemática constitui-se em um conjunto de procedimentos cujo objetivo é construir um paralelo para tentar explicar, matematicamente, os fenômenos presentes no cotidiano do ser humano, ajudando-o a fazer previsões e tomar decisões.

Metodologicamente, esta pesquisa se dividiu em 3 etapas:

Na primeira etapa foi feita uma contextualização e levantamento de um referencial teórico sobre indicadores econômicos, as aplicações financeiras e os métodos de análise de dados.

Na segunda etapa foram coletados dados obtidos sobre as variações das aplicações financeiras e dos indicadores econômicos, identificando quais serão utilizados para a modelagem matemática de análise de dados.

Terceira etapa foi realizada a análise dos resultados obtidos e a divulgação das possíveis relações entre os indicadores e aplicações financeiras.

2. REFERENCIAL TEÓRICO

2.1 Fundos de Investimentos

Os fundos de investimentos se popularizaram por oferecerem uma forma simples de investir no mercado financeiro, no qual a complexidade aumenta diariamente. Independentemente do valor aplicado, os investidores recebem o mesmo tratamento, rentabilidade e informações. Segundo Vicensi (2003, *apud* HIRT 2004, p. 19) “Investir em fundos significa acesso aos vários mercados, poder diversificar e contar com um gerenciamento profissional”. Ainda é ressaltado pelo autor, o fato de que as instituições que trabalham nesse nicho de mercado, contam com equipes de profissionais preparados, que administram as carteiras de investimento dos fundos, avaliando sempre os melhores momentos para compra ou venda dos títulos. Acrescentando nisso, quando o mercado é acompanhado com constância, é possível identificar com maior precisão as melhores oportunidades oferecidas pelo mercado.

Segundo Neto (2003, *apud* HIRT 2004, p. 20) “fundo de investimento nada mais é que um grupo de pessoas, que se conhecem ou não, que investem seu dinheiro de forma condominial.”. Uma grande parcela de investidores aplica seu capital no mercado financeiro por meio dos fundos de investimentos. Essa forma de investimento permite aplicar recursos em carteira de ativos constituída de diversos investidores, a qual será administrada por profissionais do mercado financeiro.

A avaliação do desempenho dos fundos de investimentos está centrada nos retornos ganhos por uns fundos, após considerar os riscos tomados durante o período de análise. Segundo Schutt e Caldeira (2016) o mundo dos fundos de investimento, contém uma ampla gama de estilos que diferem no que diz respeito à ativos financeiros e estratégias de trading adotadas pelos gestores quando montam e administram suas carteiras, sendo que as diferenças de estilo de investimento, ou diferentes exposições aos fatores de risco, podem explicar boa parte da variação nos retornos dos fundos de investimento.

Segundo Castello de Moraes (2000, *apud* MARTINI 2013, p.2), investir significa adiar um consumo presente, para que num futuro próximo, o investidor

possa ter mais dinheiro para consumir. De acordo com a Introdução de Mercado de Capitais da BMF&Bovespa (2010), pode-se dizer que:

Todo investidor busca a otimização de três aspectos básicos em um investimento: retorno, prazo e proteção. Ao avaliá-lo, portanto, deve estimar sua rentabilidade, liquidez e grau de risco. A rentabilidade está sempre diretamente relacionada ao risco. Cabe ao investidor definir o nível de risco que está disposto a correr em função de obter maior ou menor lucratividade. (BMF&BOVESPA, 2010: 09)

O capital do mercado de renda fixa, apesar de amplo, pode ser coberto principalmente entre 10 principais administradores, conforme Tabela 1, onde mostra os maiores administradores dos fundos de Renda Fixa, segundo ANBIMA.

Tabela 1 - Ranking de Administração de Fundos de Investimento

Ordem	Administrador	Renda Fixa
1	BB DTVM S.A	597.050,10
2	ITAU UNIBANCO SA	388.073,30
3	BRADESCO	287.737,80
4	CAIXA	282.563,30
5	BANCO SANTANDER (BRASIL) SA	173.123,20
6	BEM	90.049,40
7	BNY MELLON SERVICOS FINANCEIROS DTVM SA	51.390,90
8	BANCO J SAFRA SA	36.490,40
9	INTRAG	34.409,90
10	BTG PACTUAL	26.101,70

Fonte: ANBIMA (2019)|

Segundo Cerbasi (2008, *apud* MARTINI 2013, p.3), os investimentos em renda fixa caracterizam-se como aqueles que geram rendimentos fixos, onde a rentabilidade é previamente determinada no momento da aplicação ou no momento do resgate dela. Em comparação ao mercado de renda variável, possui ganhos expressivamente menores, com um grau de segurança bem superior. Esse tipo de aplicação é procurado pelos mais conservadores, que procuram não arriscar muito seu capital acumulado. Apesar dessa maior segurança, mesmo na renda fixa ainda existe a possibilidade de perda do capital investido, no todo ou em parte. Pode

ocorrer em casos que o emissor do título não cumpre com obrigações assumidas, neste caso o investidor deixa de receber uma parcial ou totalmente o valor investido.

Na renda fixa, o cálculo da remuneração já é previamente definido no momento da aplicação. Segundo (ANDREZO & LIMA, 2001), O mercado de renda fixa caracteriza-se pelo conhecimento do ganho futuro, em termos nominais (taxa pré ou pós-fixada), em contrapartida no mercado de renda variável, o lucro será conhecido apenas na data da venda do papel. Não obstante, devemos reconhecer que, no mercado de renda fixa, um eventual ganho nominal, considerado na compra do título, pode não se concretizar e, até mesmo, transformar-se em perda, em termos reais, devido às condições do mercado durante o período, como pode ocorrer no caso de uma elevação das taxas de juros do mercado.

Ao comprar títulos de renda fixa, sendo ele emitido pelo governo ou por empresa privadas, você está emprestando a quantia investida ao emissor do título, recebendo em retorno juros sobre o valor aplicado, como forma de remuneração. Segundo (ANDREZO & LIMA, 2001) os requisitos do investimento, como: prazos, formas de remuneração, cláusulas de recompra são acertadas com o devedor (emissor do título) no momento em que se realizar aplicação. Hoje os bancos dispõem diversas formas de aplicações de renda fixa, porém serão abordadas apenas as 10 principais.

O universo dos títulos de renda fixa incorpora os seguintes tipos:

- Prefixados: que apresentam fluxo de pagamento constante;
- Indexados à inflação;
- Com taxa flutuante: no Brasil, a taxa flutuante mais comum é a Selic, variando diariamente.

Análise da modelagem deste processo pode auxiliar na tomada de decisões para maximizar os lucros dentro dos fundos estudados e gerar um conhecimento auxiliando em investimento com outros fundos. Buscando entender de que forma os fundos de renda fixa são influenciados pelas variações dos principais índices da economia. Segundo Macedo e Macedo (2007 *apud* David 2013) fundos de renda fixa são os mais seguros e vendidos do mercado, para pessoas físicas, em comparação aos de renda variável. Realizar uma boa análise de como alguns dos principais fundos varia com base nas flutuações do mercado, vem a auxiliar na tomada de

decisões sobre quais ativos adquirir, contribuindo para independência financeira da população e o acesso a uma maior qualidade de vida.

Historicamente, a ideia de mensurar eficiência de fundos remota ao início de século 19 segundo Macedo e Macedo (2007 *apud* David 2013 p. 9), metodologias multicritérios apresentam vantagem sobre outros modelos de análise mono/bidimensionais, já que medidas de desempenho que considerem somente a relação risco *versus* retorno, ignoram informações sobre custos de transação (como a taxa de administração), ganhos de escala, experiência do gestor e outras variáveis importantes para a análise de desempenho.

Nesse sentido, Eduardo (2008, *apud* Rocha 2013) evidencia a existência de múltiplas métricas que averiguam em um só índice os atributos de risco e de retorno e que, conseqüentemente, tornam-se adequadas à análise e à comparação de performance de ativos. Os modelos mais apontados para avaliação de desempenho do retorno de uma carteira de modo a considerar o seu risco são, segundo Aragão e Barros (2004, *apud* Rocha 2013), as de Sharpe, Treynor, Sortino, Jensen e outros índices, cada uma com objetivo específico.

De acordo com Bodie, Kane e Marcus (2015, *apud* FONSECA 2017 p.10) “Investimento é o comprometimento presente de dinheiro ou outros recursos na expectativa de colher benefícios futuros”.

O mercado de fundo de investimentos está em pleno crescimento ao longo dos anos no Brasil, ainda pouco comparado aos países mais desenvolvidos, porém em franca expansão. Khorana, Servaes e Tufano (2003, *apud* HIRD 2004 p.15) analisaram o mercado de fundos mútuos em dezenas de países e concluíram que este mercado, é o preferido pelos investidores. Principalmente naqueles em que há uma legislação específica sobre o assunto, fornecendo maior segurança para aqueles que investem. Com este crescimento do número de investidores individuais nestes fundos, alguns deles vêm se sobressaindo sobre os demais como os de multimercado.

Conforme Schutt e Caldeira (2016) O universo de investimentos, devido a sua grande amplitude, é composto por os mais variáveis estilos, principalmente no que diz respeito aos ativos financeiros, objetivos de performance e estratégias de trading adotadas pelos gestores ao montarem suas carteiras, sendo que as diferenças de

estilo de investimento, ou diferentes exposições aos fatores de risco, podem explicar boa parte da variação *cross-sectional* dos fundos de investimento.

Segundo dados da Anbima (2012), apenas no período de 2002 a 2012, houve um incremento de mais de 200% nas quantias administradas por fundos de investimentos no Brasil. Conforme Barros e Amaral (2003 *apud* David 2013 p.8) a avaliação de carteiras de investimento passou a ser um assunto em expansão na literatura brasileira.

De acordo com Fonseca (2017), uma das maneiras mais comum, de se medir o desempenho do investimento é comparando com índices apropriados de mercado. Eles fazem parte do cotidiano tanto do especialista em investimentos, quanto do indivíduo comum, que usufrui apenas dos conhecimentos advindos de noticiários referentes às flutuações de mercado.

Nesse sentido, Macedo e Macedo (2007, *apud* ROCHA 2013 p. 8) relatam que os critérios dessa avaliação:

Abordam questões que justificam a melhor ou a pior performance dos gestores de fundos. O sucesso na administração de carteiras de investimento costuma ser atribuído a esses diversos fatores, que, de modo geral, podem ser encontrados por meio da mensuração de algumas variáveis. Por isso, avaliar o desempenho dos fundos de investimentos é o melhor caminho para se identificar alguma falha na utilização desses fatores na gestão das carteiras.

Conforme Fonseca (2017), os índices financeiros constantemente rodeiam o âmbito acadêmico. De acordo com o modelo de índice único de Sharpe (1963, *apud* Fonseca 2017 p.10):

Variação do preço de uma ativo é uma função da variação da carteira de risco do mercado, sendo o índice financeiro uma proxy para uma carteira hipotética que contém todos os ativos de risco do mercado.

Fundos de investimentos indexados mostraram nos últimos 20 anos crescimento robustos, obtendo uma fatia substancial do montante de recursos investidos. Em 2015, a totalidade dos ativos destes fundos nos Estados Unidos era cerca de US\$4 trilhões, valor proporcional a 34% do total investidos em fundos de investimento, já em 1995 essa soma era de apenas 4% (BOGLE, 2016).

Ainda segundo Bogle (2016), tarifas para gerenciamento reduzidas, incompetência de gestores adeptos a técnica de superar os índices e aumento da instabilidade com a crise estadunidense de 2008, são os elementos imprescindíveis para esse crescimento substancial.

Segundo Silva (2016), nos países com mercados mais avançados, a curva de rendimento dos títulos prefixados, é tomada como pilar de parâmetro para todos os outros títulos de renda fixa, como nos casos dos membros da zona do Euro e da economia americana. O mercado de títulos públicos líquidos aprimora e potencializa o mercado futuro.

Segundo Schutt e Caldeira (2016), investidores pragmáticos procuram atingir uma proporção permissível dentro das circunstâncias de risco e retorno. Nessa razão, a destinação estratégica de ativos é de vasta relevância na gestão de portfólios, e a instalação de um equilíbrio pende de uma combinação competente de ativos. Entretanto, com certa assiduidade, muitos investidores, e mesmo a imprensa conhecedora do assunto, confrontam unicamente os retornos obtidos de fundos de investimento sem levar em conta as dessemelhanças na sua exposição. Em função da larga quantidade de fundos presentes hoje em dia, se torna muito difícil para o investidor identificar de uma maneira adequada qual tipo de investimento lhe convém, e quais fundos lhe atendem mais especificamente. De tal forma, para que o investidor possua as informações fundamentais para estabelecer qual fundo é mais apropriado ao seu perfil de risco e retorno é relevante gozar de ferramentas que permitam distinguir as particularidades e fatores de risco de cada fundo.

2.2 Índice Nacional de Preços ao Consumidor Amplo

A inflação é um assunto que por toda extensão da história econômica brasileira causou e continua causando preocupação a população. No Brasil, a inflação começa devido a uma rápida amplificação de crédito para o público em geral. Economistas e historiadores relatam o primórdio da explosão inflacionária no final do século XIX, o qual deu início a um ganho maior de destaque durante a década de 1980 (CARDOSO, 2007, *apud* GONÇALVES, *et al.*, 2017).

Nos últimos 25 anos, essencialmente após a implantação do Regime Monetário de Metas de Inflação em 1999, as taxas de inflação têm se apresentado

mais uniformes, sem grandes variações, pelo menos se confrontadas com o árduo período pelo qual o Brasil percorreu desde a década de 1980 até a implantação do Plano Real em 1994.

Desde tal implantação, estabeleceu-se claramente como uma preocupação que boa parte do foco da política econômica deveria ser o atingimento das metas estabelecidas (LEITE; ALMEIDA, 2012, *apud* GONÇALVES *et al.*, 2017). Porém, a partir de 2010, percebe-se certas adversidades em reter a inflação em níveis razoavelmente próximos aos estipulados pelos formuladores de política Bacen (2016, *apud* CARRARA; BARROS, 2016). Tal dificuldade dá abertura para discussões e estimula estudos que tentam descobrir suas possíveis causas.

O IPCA (Índice Nacional de Preços ao Consumidor Amplo) é o índice mais importante do ponto de vista da política econômica. Além de ser, segundo *Oliveira et al.* (2011 *apud* GONÇALVES *et al.*, 2017), medidor legal da inflação no Brasil, foi definido como parâmetro destinado ao sistema de metas de inflação instalado no Brasil em 1999 pelo Conselho Monetário Nacional.

O IPCA verifica as mudanças de valores nos produtos consumidos por famílias com renda de 1 a 40 salários-mínimos, em nove regiões metropolitanas (Belém, Belo Horizonte, Curitiba, Fortaleza, Porto Alegre, Recife, Rio de Janeiro, Salvador e São Paulo), no Distrito Federal e no município de Goiânia Mendonça (2007, *apud* GONÇALVES *et al.*, 2017). Também segundo Mendonça:

De acordo com Barbosa(2018), o IPCA é um índice de Laspeyres que calcula variação dos preços de bens e serviços entre dois períodos (comparação dos preços vigentes nos trinta dias do período de referência e com os trinta dias do período base), ponderando a participação dos gastos com cada bem no consumo total.

Segundo Barbosa (2018), os índices de inflação podem ser estabelecidos baseados em diversas variáveis de cestas de bens, de acordo com os preços significativos para os grupos que se queira conceituar: produtores, consumidores de um setor, ou da economia em geral, residentes de uma determinada região, dentre outros. Se analisarmos os modelos macroeconômicos, veremos que frequentemente tratam a inflação como uma única medida, procurando quais são seus determinantes e as políticas mais apropriadas para controlá-la.

Autores como Mallick e Sousa (2013, *apud* CARRARA; BARROS, 2016), entre outros, produziram estudos, em diversos países e avaliando épocas diferentes, na busca de apontar a origem da variação da inflação, e chegaram a uma mesma resposta, que essas alterações, se devem parcialmente aos impactos causados pelas ofertas. No caso da economia brasileira, Moreira (2014, *apud* CARRARA BARROS, 2016) aponta que “os choques de oferta, ligados aos preços das commodities, afetam de forma importante as expectativas de inflação”.

A taxa de inflação, traz uma visão mensurada do crescimento generalizado dos preços de artigos básicos que compõem a economia nacional. Segundo Vilaça (2015):

Os modelos ortodoxos usualmente retratam a inflação como grandeza única, buscando seus determinantes e as políticas adequadas para contê-la. A corrente teórica é a base subjacente à utilização dos instrumentos de política econômica.

De acordo com Resende & Pelicioni (2009, *apud* Vilaça 2015), para os seguidores dessa visão ortodoxa, se vale a visão clássica que, ao menos no longo prazo a moeda é isenta, não tendo influências diretas no lado “real” da economia. Já para a visão heterodoxa, o ambiente financeiro age em conjunto com universo real, determinando emprego, os preços e produto, assim sendo, verifica-se que a moeda não é neutra, tendo interação direta sobre as decisões de consumo e o acúmulo de capital.

Segundo dados do IPEA, a inflação brasileira teve um processo de aceleração iniciado, por volta de 2010, assim sendo, suas taxas de variação acumuladas nos períodos analisados de 12 meses, se encontram não apenas acima da meta de 4,5%, como também além da medida superior de tolerância estabelecida de 6,5%.

2.3 Taxa de Câmbio

A taxa de câmbio é um dos preços relativos básicos das economias de mercado, tem tamanha importância devido ao fato de intermediar as economias nacionais com o resto do mercado global. Existem modelos analíticos que geram um equilíbrio geral, permitindo definir que a taxa de câmbio se adequa quando transmite

a competitividade externa do país, bem como a confiança sobre fatores macroeconômicos essenciais.

O regime cambial do Brasil é de monopólio cambial, ou seja, só o Banco Central pode efetuar legalmente transações com divisas estrangeiras no país ou autorizar que agentes as façam, sob sua fiscalização. Taxa de câmbio, por sua vez, não deve ser confundida com monopólio cambial, a mesma é estabelecida pelas forças do mercado e pelas intervenções do banco Central. Os controles do mercado cambial estão estabelecidos nas Leis 4 131/62 e 4 595/67.

Segundo BRESSER-PEREIRA (2012, p.10) uma boa taxa de câmbio é essencial para o desenvolvimento econômico pois funciona como um interruptor de luz que “ativa” ou “desativa” as empresas tecnológica e administrativamente competentes à demanda mundial. Uma boa taxa de câmbio estimula os investimentos voltados para a exportação e aumenta correspondentemente a poupança interna.

A taxa de câmbio é uma variável importante para as economias mundiais. Segundo (DOS SANTOS, 2013) a mesma pode ser determinada como fixa ou flutuante, a taxa de câmbio é o preço de papel estrangeiro em comparativo com uma moeda nacional do país estudado. O mercado cambial pode ser analisado através das diversas moedas existentes no planeta, sendo o responsável por suportar as transações financeiras, de capital ou correntes (importações e exportações), dentre as economias do globo a fora, porém a principal moeda estrangeira utilizada como parâmetro é o dólar norte-americano devido a sua maior liquidez.

Segundo Nonnenberg (*et al.*, 2015) o conceito de taxa efetiva real de câmbio é a unificação, em uma única variável, de três prerrogativas fundamentais para a compreensão da competitividade de um país com os outros: a relação entre os preços domésticos e externos, que se reflete nessa taxa.

Uma alta na taxa de câmbio se traduz como necessário fazer uso de mais unidades da moeda doméstica para a compra da moeda estrangeira utilizada como parâmetro, isto é, há uma depreciação da moeda doméstica. Todavia, uma baixa na taxa de câmbio se traduz em utilizar menos unidades monetárias necessárias para a compra de uma unidade de moeda estrangeira, ou seja, uma valorização da moeda nacional em comparação as do exterior. Uma taxa de câmbio considerada

competitiva é aquela que corresponde à taxa de câmbio necessária para que empresas que utilizem tecnologia em nível global sejam internacionalmente competitivas com suas concorrentes internacionais de mesmo nível e âmbito.

Segundo Nonnenberg (*et al.*, 2015) a relação que prevalece no mundo é entre a moeda nacional e dos outros países com os quais o país estabelece uma relação. Não se restringindo somente a transações de bens, mas também serviços, e investimentos.

Pode-se verificar que são inúmeras as variáveis econômicas empregues por modelos econômicos tentando antecipar a taxa de câmbio, como, PIB, oferta monetária, taxas de juros, nível de preços, entre outras. No entanto, segundo Dos Santos (2013) o que se institui são variantes nos modelos para o curto e longo prazo. De modo geral, é compreensível que fundamentos macroeconômicos possam estabelecer a taxa de câmbio no longo prazo e isso é aceito pelos profissionais da área e imprensa especializada. Porém no curto prazo a explanação para mudanças do câmbio torna-se um trabalho árduo de se estabelecer, até mesmo após uma variação bem elevada, como no caso do Brasil, em que o real já chegou a desvalorizar mais de 8% apenas em um dia.

2.4 Taxa de Juros (CDI)

As taxas de juros constituem uma das variáveis macroeconômicas mais fundamentais para o bom funcionamento da economia. Segundo (NETO, 2019). uma taxa de juros bem calibrada é essencial, pois os juros têm papel fundamental na determinação do nível de atividade, emprego, taxa de câmbio e outras variáveis. No Brasil, é comum encontrarem-se queixas quanto ao alto nível das taxas de juros, porém no passado, juros baixos foram apontados por muitos como os culpados pela nossa hiperinflação.

No Brasil, diferentemente do que ocorre em diversos países, os títulos são transacionados no mercado por uma taxa de juros, ou um spread de taxa de juros. Spread seria explicado como um adicional de taxa, cobrado sobre uma taxa de referência. Segundo (SILVA, *et. al.*, 2016) “Taxa CDI é uma média dos juros praticados pelos Certificados de Depósitos Interbancários realizados entre os

Bancos.” Esta taxa também é chamada de Taxa DI (Taxa dos depósitos Interbancários), ambas se referem a uma mesma taxa.

O CDI-Certificado de Depósito Interbancário, é um título emitido com a intenção de captação de capital, para as instituições financeiras, seja bancos, financiadoras, etc. Segundo Silva (*et al.*, 2016) o CDI difere de outros ativos, pois a transação é efetuada exclusivamente de uma instituição financeira para outra. Em ocasiões que se julgue necessário um banco tomar dinheiro emprestado de um segundo Banco, ele é capaz de emitir um CDI e recolher este dinheiro no Mercado Interbancário

O CDI tem cotações muito próximas da Taxa Selic, muitas vezes sendo confundidas, isso ocorre para evitar arbitragens, no mercado interbancário, porém não será o foco a correlação de ambos neste trabalho. A taxa CDI é determinada pela troca de recursos realizada entre os bancos. Segundo (SILVA *et al.*, 2016).

O CDI é muito utilizado para os investimentos de renda fixa porque a taxa CDI é a média pela qual os Bancos captam recursos, e em muitos momentos, é também a taxa de juros mais competitiva que eles conseguem para captação de recursos. Sendo assim, a taxa que os bancos remunerarão os investimentos dos seus clientes, ou outra captação que consigam sempre vai estar relacionada a taxa CDI, seja como % do CDI, seja tentando superar o CDI. Muitos fundos de investimento cobram a taxa de performance baseados no quanto o fundo ultrapassa a taxa CDI.

Esta taxa na realidade chama-se apenas Taxa DI (Taxa dos depósitos Interbancários), entretanto, seja no meio profissional ou noticiado diariamente, as pessoas a tratam e conhecem por Taxa CDI, ambas representam uma mesma taxa. De tal forma como a Taxa Selic, que é empregada para mensurar os juros dos títulos públicos, e é calculada como uma média conseguida pelo sistema Selic; a Taxa CDI é calculada através do sistema Cetip (SILVA, *et al.*, 2016)

O CDI é uma referência amplamente utilizada para verificar se um investimento é bom ou não. Segundo Silva (*et al.*, 2016) ele é muito empregado pelos investimentos de renda fixa porque a taxa CDI é a média pela qual os Bancos captam recursos, e em diversas ocasiões, é inclusive a taxa de juros mais

competitiva que eles conseguem para obtenção de capital. Sendo assim, a taxa que os bancos recompensarão os investimentos de sua clientela, ou outra captação que consigam, permanentemente vai estar interligada a taxa CDI, seja como % do CDI, seja tentando superar o CDI.

Diversos fundos de investimento exigem a taxa de desempenho baseados no quanto o fundo sobressai em relação a taxa CDI. Períodos de alta dos juros com a finalidade de segurar a inflação podem gerar danos na rentabilidade das carteiras de renda fixa. Isso acontece, devido ao fato do valor de mercado dos títulos já emitidos, cair, provocando prejuízo para os detentores dos papéis (COSTA, 2013, apud SILVA et al., 2016)

No período de três anos (jan. de 2013 a jan. de 2016), a taxa básica de juros variou entre a mínima de 7,25% ao ano e a máxima de 14,25% ao ano.

3. MODELOS DE ANÁLISE DE DADOS PARA SÉRIES ECONÔMICAS

3.1 Classificação

3.1.1 Aprendizagem Bayesiana

A aprendizagem Bayesiana está fundamentada na presunção de que as grandezas de interesse são reguladas por distribuições de probabilidades. Tendo que mensurar o custo/benefício entre decisões divergentes de classificação utilizando probabilidades e custos afiliados à classificação.

Redes Bayesianas, similarmente divulgadas como Teoria da Crença, surgiu como um método probabilístico que concede a representação eficiente do conhecimento para previsão ou assimilação de um certo estado (SALAZAR, 2017). Para Russell e Norvig (2013, apud CAMPOS et al., 2019) “a análise preditiva das redes bayesianas baseia-se em conjuntos rotulados e prévios de informação histórica”. Segundo Melrose (et al., 2015, apud SALAZAR, 2017), as teorias Bayesianas são utilizadas em larga escala no âmbito da inteligência artificial, para tratamento de incertezas e auxílio na tomada de decisões sistêmicas.

Tais redes integram um modelo gráfico que demonstra simplificada as relações de causalidade, de determinadas variáveis de um sistema. Em decorrência desse fator, ele é amplamente utilizado para medir riscos operacionais, cálculo de sensibilidade de perdas em eventos, detecção de padrões e tendências (MARQUES; DUTRA, 2002, *apud* SALAZAR, 2017).

O algoritmo de classificação Naive Bayes, determina a probabilidade de um item fazer parte de uma classe estipulada, considerando que suas características sejam condicionalmente independentes (TAN; STEIBACH; KUMAR, 2009, *apud* DIAS, 2018).

Este algoritmo supõe a independência condicional entre itens e classes, suposição essa que obviamente difere da prática real, pois os atributos em conjuntos reais de dados estão em boa parcela correlacionados ou inclusive limitado a uma classe específica. Mesmo com essa aproximação, o algoritmo de classificação Naive Bayes consegue entregar bons resultados em diversos domínios (AGGARWAL, 2015, tradução nossa, *apud* DIAS, 2018).

3.1.2 Por árvores de decisão

Árvore de decisão ou de classificação é um método que têm o objetivo de prever um desfecho a partir de variáveis preditoras. Estas árvores são elementos gráficos utilizados para calcular as probabilidades de determinada situação ocorrer, considerando a presença/ausência de certos conjuntos de fatores. Estes elementos gráficos permitem uma visão global de múltiplas combinações de status, com uma medida de probabilidade para todos os status (CHAVES *et al.*, 2018). Árvores de decisão é um exemplo de modelo estatístico que usa classificação e predição de dados, além de ser um dos algoritmos mais usados para predição (SANTOS *et al.*, 2015).

3.1.3 Por regras

De acordo com Skinner (1969, *apud* PINTO, *et al.*, 2006),

“Regras são estímulos especificadores de contingências e exercem controle como estímulos discriminativos, fazendo parte de um conjunto de contingências de reforço”.

Além disso, verifica-se por meio de estudos realizados, que a forma por meio da qual o procedimento é determinado (contingências ou por regra), é capaz de afetar na adaptação do comportamento a mudanças nas contingências (PARACAMPO, *et al.*, 2001, *apud* PINTO, *et al.*, 2006).

Segundo Lowe (1979, *apud* PINTO, *et al.*, 2006), humanos mesmo quando não estão sendo expostos a regras, podem apresentar suas próprias regras. Conforme mostrado por Ayllon & Azrin (1964, *apud* PINTO *et al.*, 2006), regras expostas pelo condutor de um estudo, têm a capacidade de causar um elevado índice autoridade sobre o comportamento humano.

3.1.4 Instâncias

No desenvolvimento de softwares, regularmente se faz necessário validar a especificação dos dados do sistema, isto é, a informação que o software controla. De acordo com Árias (2011) os dados trabalhados e armazenados por aplicações são comumente descritos por esquemas. O esquema de dados é quem determina disposição lógica e as relações entre os dados.

A análise por instâncias de dados, pode seguir pela forma da AIDA (Análise de Instâncias de Dados Alternativas), neste formato de análise, uma instância de dados agregada ao esquema em teste sofre alterações de forma simplificada, constituindo instâncias de dados alternativas. Segundo (ÁRIAS, 2011) Cada uma dessas alternativas simboliza possíveis defeitos que podem estar presentes no esquema e é formada a partir de classes de defeitos previamente estabelecidas com início em um modelo genérico capaz de representar os mais variados tipos de esquemas de dados. Consulta às instâncias geradas são realizadas e a consequência dessas consultas, caso difira do resultado aguardado, aponta a existência de defeitos no esquema, pois a instância consultada foi considerada como válida para esse esquema.

Essa abordagem pode ser aplicada em testes de diferentes formatos de esquema, desde que tal esquema seja representado de acordo com o modelo formal.

3.1.5 Regressão logística

Segundo Lopes *et al.* (2017), o modelo de regressão logística, está sendo considerado o mais utilizado no mercado atual. Constitui-se numa técnica de análise multivariada com intuito de checar as ocorrências de elemento e eventos referentes às variáveis. Essa técnica utiliza ensaios estatísticos diretos, perícia de incorporar efeitos não lineares e diversos tipos de diagnósticos.

Esta técnica é preferida à análise discriminante por ser similar à regressão, com testes estatísticos diretos, habilidade de incorporar efeitos não-lineares e diversos tipos de diagnósticos. Na estimação dos coeficientes da função logística, esta técnica procura maximizar a verossimilhança de que um evento ocorra de acordo com HAIR JR *et al.* (2009, *apud* GOUVÊA, 2012).

A regressão logística é uma técnica estatística multivariada que já vem sendo amplamente utilizada no setor de crédito bancário, por ter eficiência na previsão dos bons e maus tomadores, analisando-os e classificando com notas posteriormente denominadas *credit scoring*. Ela é considerada o método de pontuação de crédito mais utilizado pelas IFs para avaliar o risco FALANGIS (2008, *apud* CAMARGOS *et.al.*, 2011).

3.1.6 Máquinas de vetores de suporte (MVS)

As Máquinas de vetores de suporte compõem uma técnica de aprendizado de máquinas com fundamentos na teoria de aprendizado estatístico de acordo com VAPNIK (1995, *apud* LORENA, 2006). Nas questões de classificação a principal finalidade é construir um preditor ou classificador, que a partir de um englobado de dados conhecidos seja capaz de indicar corretamente a classe de novos exemplos. Essas são visualizadas como categorias dos dados, e são capazes de descrever a ocorrência de interesse sobre o qual é desejado realizar previsões. (Monard & Baranauskas, 2003 *apud* LORENA, 2006).

As MVS, foram originalmente desenhadas para a discriminação de dados participantes a duas classes, ou assim dizendo, de classificação binária. Entretanto diversas situações práticas possuem mais que duas classes preditivas, sendo tratadas através do termo multiclass. Esse problema pode ser considerado diversas vezes mais complexo que um problema binário, dado ao fato que o modelo

de predição gerado deve ter a capacidade de dividir os dados em um número elevado de categorias, aumentando assim o percentual de erros de classificação ocorrerem. De tal forma, sua complexidade é diretamente proporcional com o número de classes elencadas Lorena, (2006).

Pode-se trabalhar com duas abordagens para a generalização das MVS à problema multiclases. De acordo com Lorena, (2006), inicialmente recomenda-se decompor o problema em diversas categorias de problemas menores binários. Sendo usadas na indução de classificadores para cada um desses novos problemas. O resultado desses classificadores combina-se na obtenção de previsão multiclases. A segunda abordagem atua alterando o algoritmo original de treinamento das máquinas de vetores suporte, dando origem à diversas multiclases. Segundo Hsu & Lin (2002, apud LORENA, 2006), tais alterações levam ao custo elevado desses algoritmos, devido a isso se torna mais comum a decomposição do problema multiclases em problemas menores binários, chamado de estratégia decomposicional.

3.1.7 Redes neurais artificiais

Redes neurais artificiais, ou também chamadas como RNA, possuem dois pontos cruciais, primeiro a arquitetura e em segundo o algoritmo de aprendizagem Rauber (2005). Tal separação surge de forma natural pelo modelo como a rede é treinada. Diferentemente de um computador com arquitetura de von Neumann que é programado, a rede é capacitada através de exemplos de treino.

De acordo com Rauber (2005) a noção sobre o problema em consideração está armazenada dentro dos exemplos que devem estar necessariamente disponíveis para checagem. É realizada uma generalização desses dados pelo algoritmo de aprendizagem e em sequência memorizado o conhecimento através dos parâmetros adaptáveis da rede, os pesos. De tal forma, o desenvolvedor do sistema baseado em redes neurais artificiais tem dois graus de liberdade, o algoritmo para capacitar e adaptar os pesos da rede e a escolha sobre qual tipo de rede para resolver o problema em questão.

A rede é constituída pelos neurônios, usualmente a forma de processamento de um único neurônio é a combinação linear das entradas com os pesos

acompanhado pela passagem da combinação linear por uma chave de ativação. Restrições normalmente são definidas através da estrutura do problema a ser solucionado. Em sua grande maioria, as implementações de RNA são realizadas em software, existindo diversas séries de simuladores publicamente disponíveis, com código fonte e manual. Após ser definida a rede neural, a mesma deve ser capacitada. Devido a isso, para que seja possível solucionar as tarefas em questão, os graus de liberdade devem ser adequados de uma maneira ótima Rauber (2005).

3.2 Regressão Linear

A regressão linear é uma das formas de modelar uma equação matemática linear que retrate o relacionamento entre duas variáveis (CURRAL, 1994, *apud* RODRIGUES *et al.*, 2013).

Análise de regressão analisa a relação entre uma variável dependente e outras variáveis chamadas independentes. Segundo (CURRAL, 1994, *apud* RODRIGUES *et al.*, 2013), esta relação é mostrada por um modelo matemático, isto é, por uma equação que associa a variável dependente com as variáveis independentes. Este modelo é designado por modelo de regressão linear simples em que é definido uma relação linear entre a variável dependente e uma variável independente. No caso de serem incorporadas várias variáveis independentes, o modelo passa a denominar-se modelo de regressão linear múltipla.

3.2.1 Não linear

De acordo com (ZEVIANI, *et. al.*, 2013) modelos não lineares (MNL) comumente são amparados por alguma informação sobre a relação entre Y e X. Tal informação está relacionada à diversos níveis de conhecimento como, análises de diagramas de dispersão, verificando y contra x; restrições de aspectos da função (ser monótona, ser sigmóide); a solução de uma equação diferencial embasada por determinado princípio ou teoria; ou a interpretação dos seus critérios. Não importa qual seja o nível de conhecimento, a opção por um modelo não linear raramente é empírica.

Os modelos de regressão assumem, de modo geral, a seguinte forma:

$$Y=f(X, \theta)+\varepsilon=E(Y|X)+\varepsilon$$

Os modelos não lineares segundo (ZEVIANI, *et. al.*, 2013) são caracterizados por funções $f(X,\theta)$ não lineares em pelo menos um dos parâmetros θ_i . Ou seja, para que determinado modelo seja não linear é preciso que ao menos uma das derivadas da $f(x, \theta)$ com relação a θ_i dependa de ao menos um dos parâmetros θ_i .

3.3 Agrupamento

3.3.1 K-means

K-means, é um algoritmo altamente utilizado para agrupamento de dados. Ele é usualmente simples de se usar, e trabalha muito bem na prática, sendo base de diversos outros métodos de agrupamento (SAMMA e SALAM, 2009, *apud* BORGES 2010). É um método de particionamento (método não-hierárquico) que divide as observações dos dados em k clusters mutuamente exclusivos.

Esse algoritmo verifica considera como centro de um grupo o seu centróide. O centróide de um grupo é determinado como o vetor soma de todos os vetores correspondentes aos objetos associados a este grupo. Logo, “o trabalho do algoritmo K-Means é minimizar a função objetivo correspondente à distância total entre os objetos e os centróides dos grupos aos quais esses objetos foram associados”. (VARGAS, *et. al.* 2011).

3.3.2 Hierárquico

Métodos de agrupamento hierárquico é um método promissor na colaboração da construção de análises exploratórias de tópicos emergentes de forma local contextualizada. Essas técnicas dividem os tópicos em grupos e subgrupos, no qual tópicos de um mesmo tema, ou similar, estão posicionados no mesmo grupo, porém sendo dissimilares aos tópicos de outros grupos. Desta maneira, o usuário pode explorar a coleção textual de forma intuitiva e em diversos níveis de granularidade (PANAGGIO, *et. al.* 2011)

Os métodos hierárquicos geram diversos grupos aninhados que são estruturados em uma árvore. Cada vértice (grupo) na árvore (exceto os vértices

folha) é a união de seus filhos (subgrupos), e a raiz da árvore é o grupo abrangendo todos os objetos (TAN *et al.* 2005, *apud* MURRUGARRA-LLERENA *et al.*, 2011).

3.4 Associação

3.4.1 A priori

De acordo com Silva *et al.* (2014) uma Regra de Associação de dados forma-se por uma expressão de implicação, representada como $X \rightarrow Y$. Tal associação, X e Y, antecedente e conseqüente, respectivamente são grupos apartados de itens. A descoberta de um padrão de implicação ocorre de existir pelo acaso, devido à isso, é necessário medir a veracidade da Regra de Associação descoberta. Segundo Tan *et al.* (2009, *apud* SILVA, 2014), essa averiguação efetuada através dos seguintes itens:

- Suporte: responsável por definir a frequência na qual uma regra é cabível a um determinado conjunto de dados. É um parâmetro essencial, pois através dele consegue-se eliminar regras sem interesse e também verificar a existência de regras que tenham reduzido suporte por coincidência.

- Confiança: responsável por definir a periodicidade na qual os itens em Y aparecem em operações que englobem X. De tal forma é possível medir a credibilidade da inferência realizada por uma regra. Exemplificando, caso tenha-se uma regra $X \rightarrow Y$, o nível de confiança determina a probabilidade, sendo assim, quanto maior a confiança, maior a probabilidade de Y estar presente em operações que englobem X.

O algoritmo A priori tem como base o princípio de que seja qual for o subconjunto de itens frequentes, o mesmo deve ser um *itemset*. Segundo Silva *et al.* (2014), fazendo uso de uma tática de busca em largura em conjunto com um algoritmo de geração e teste. Em cada grau são produzidos os *itemsets* viáveis, considerando os mais frequentes gerados no grau anterior, posteriormente a geração deles, é testado a regularidade, atravessando por completo novamente a base de dados de transações.

De acordo com Silva *et al.* (2014), o algoritmo efetua uma pesquisa por largura, podendo ser definido em três atributos: K-itemsets, suporte mínimo e

confiança mínima. Se escolhe o valor mínimo do suporte para que um K-itemset seja considerado e a confiança limita a filtragem das associações recolhidas pelo algoritmo.

3.4.2 ECLAT

O algoritmo ECLAT, também conhecido como *Equivalence Class Transformation*, no modelo paralelo proposto por Zaki *et. al.* (1997, *apud* MARIANO, 2011) baseia-se em um método que se desdobra todo o espaço de busca inicial em diversas subdivisões menores, de modo que cada subdivisão possa ser processada independentemente. Esses desdobramentos são realizados dissociando o espaço de busca aplicando a técnica de clusterização por classes de equivalência, como utilizado no modelo sequencial.

De acordo com Mariano (2011), no algoritmo, o local de procura refere-se ao banco de dados a ser processada, na qual é compartilhada entre os processadores, de modo que cada processador seja capaz de descobrir os *itemsets* frequentes independentemente, empregando intersecções de *tidlists*. Para esse fim, uma projeção vertical do banco de dados deve ser utilizada, em que é erguida com apenas uma varredura no banco de dados.

No modelo paralelo, é consistido por duas fases, fase de inicialização e fase de processamento. Na primeira, a técnica de clusterização por classes de equivalência é utilizada distribuindo as classes definidas entre os processadores. Na segunda fase, de processamento, desde as classes de equivalência, cada processador define seu grupo de *itemsets* constantes independentemente.

Já utilizando uma busca em profundidade, segundo Mariano (2011), o algoritmo ECLAT, emprega um método que combina a procura em profundidade com intersecções entre conjuntos. Nessas procuras em profundidade, se faz necessário armazenar em memória as *tidlists* dos *itemsets* correntemente em análise.

O algoritmo ECLAT emprega uma otimização nomeada como “intersecções rápidas”. Na execução de uma intersecção entre duas *tidlists*, a *tidlist* derivada tem em sua composição os *itemsets* que contém suporte maiores que o mínimo

especificado *min_sup*. Nas chamadas “intersecções rápidas”, a técnica de intersecção é suspensa no momento em que é realizado não ser possível atingir o suporte mínimo Mariano (2011).

4. DESENVOLVIMENTO DO TRABALHO

4.1 WEKA como suporte a análise de dados

Segundo a UNIVERSIDADE DE WAIKATO (2021), o WEKA (*Waikato Environment for Knowledge Analysis*) é um programa o qual já possui uma biblioteca de algoritmos de Classificação e *Machine Learning* pré-instalada, com o intuito principal de trabalhar com as tarefas de mineração de dados. Ele possui desde ferramentas utilizadas na preparação, classificação, quanto na parte de regressão, agrupamento e por visualização de dados. Sendo largamente empregado na área de educação e pesquisa, com seu elevado número de instrumentos integrados para tarefas de aprendizado de máquina padrão. Os dados dos fundos e indicadores foram extraídos dos sites de organizações do governo e empresas responsáveis pela gestão dos fundos, no formato CSV, configuração essa que é aceita pelo WEKA.

4.2 Definição dos fundos a serem investigados

Os fundos foram definidos com base no ranking de maiores participações de mercado no estoque de crédito, analisando apenas Pessoas físicas, com bases em informações extraídas do Relatório de Economia Bancária de 2020 do Banco Central. Trazendo um fundo de cada um dos dois maiores bancos em 2020 nesse quesito, conforme Figura 1 abaixo, onde podemos avaliar os maiores bancos em participação de mercado no estoque de crédito.

Figura 1 - Participação de mercado no estoque de crédito - Pessoas físicas

Posição	2018	2019	2020	%		
1	Caixa Econômica Federal	29,4	Caixa Econômica Federal	27,3	Caixa Econômica Federal	27,0
2	Banco do Brasil S.A.	18,6	Banco do Brasil S.A.	18,0	Banco do Brasil S.A.	17,4
3	Itaú Unibanco S.A.	12,0	Itaú Unibanco S.A.	12,2	Itaú Unibanco S.A.	11,9
4	Banco Bradesco S.A.	10,5	Banco Bradesco S.A.	11,3	Banco Bradesco S.A.	11,2
5	Banco Santander (Brasil) S.A.	9,6	Banco Santander (Brasil) S.A.	10,0	Banco Santander (Brasil) S.A.	10,0
6	Demais bancos comerciais e múltiplos com carteira comercial	11,5	Demais bancos comerciais e múltiplos com carteira comercial	12,0	Demais bancos comerciais e múltiplos com carteira comercial	12,7
7	Cooperativas de crédito	4,1	Cooperativas de crédito	4,5	Cooperativas de crédito	5,2
8	Bancos de investimento/ múltiplos sem cart. comercial	2,3	Bancos de investimento/ múltiplos sem cart. comercial	2,4	Bancos de investimento/ múltiplos sem cart. comercial	2,3
9	Segmento não bancário	1,7	Segmento não bancário	2,1	Segmento não bancário	2,1
10	Bancos de desenvolvimento	0,2	Bancos de desenvolvimento	0,2	Bancos de desenvolvimento	0,0
Total		100,0		100,0		100,0
IHHn		0,1579		0,1455		0,1414
Equiv do IHHn		6,3		6,9		7,1
RC5(%)		80,1		78,7		77,6

Fonte: Relatório de Economia Bancária, 2020

O quesito a ser analisado de cada fundo é a rentabilidade dele. Define-se a rentabilidade como o retorno bruto do fundo, descontado de todos os custos e taxas, mas sem considerar os impostos. Será analisado um período de 36 meses, considerando os anos fechados de 2018, 2019 e 2020. Tendo desta forma uma base de retorno mensal para os fundos.

O primeiro fundo que será analisado é o FUNDO INV EM COTAS FUNDOS INV CAIXA GERAÇÃO JOVEM RENDA FIXA CRÉDITO PRIVADO LONGO PRAZO (CNPJ: 10.577.485/0001-34) da Caixa Econômica Federal. O público-alvo deste fundo segundo a organização são investidores que buscam retorno por meio de investimentos em fundos de renda fixa e o fundo destina-se a acolher investimentos de pessoas físicas e jurídicas.

O fundo tem como objetivo gerar aos seus investidores rentabilidade por meio de aplicação em cotas de Fundos de Investimento os quais façam uso de aplicações em carteira diversificada de ativos financeiros de renda fixa.

A Figura 2 apresenta as condições necessárias para realizar investimento no fundo, segundo a Caixa Econômica Federal.

Figura 2 - Condições de Investimento – CAIXA ECONÔMICA FEDERAL

Investimento inicial mínimo	R\$ 0,01
Investimento adicional mínimo	R\$ 0,01
Resgate mínimo	R\$ 0,01
Horário para aplicação e resgate	17:00 (horário de Brasília)
Valor mínimo para permanência	R\$ 0,01
Período de carência	Não há
Conversão das cotas	Na aplicação, o número de cotas compradas será calculado de acordo com o valor das cotas no fechamento da data da aplicação. No resgate, o número de cotas canceladas será calculado de acordo com o valor das cotas no fechamento da data do pedido de resgate.
Pagamento dos resgates	O prazo para o efetivo pagamento dos resgates é de 0 dias úteis contados da data do pedido de resgate.
Taxa de administração	1,00% ao ano
Taxa de entrada	Não há
Taxa de saída	Não há
Taxa de performance	Não há
Taxa total de despesas	As despesas pagas pelo fundo representaram 1,27% do seu patrimônio líquido diário médio no período que vai de 01/06/2020 à 31/05/2021. A taxa de despesas pode variar de período para período e reduz a rentabilidade do fundo. O quadro com a descrição das despesas do fundo pode ser encontrado em www.caixa.gov.br

Fonte: CAIXA ECONOMICA FEDERAL (2021).

Os dados históricos de rentabilidade do fundo foram extraídos do próprio site da Caixa Econômica, detentora do fundo a ser estudado.

O segundo fundo a ser analisado é o fundo BB BESC RENDA FIXA PRÁTICO CRÉDITO PRIVADO FUNDO DE INVESTIMENTO (CNPJ: 00.073.041/0001-08) do Banco do Brasil. Segundo a companhia, o fundo é destinado a investidores que pretendam buscar retorno por meio de investimentos em ativos de renda fixa, podendo manter mais de 20% da sua carteira em títulos de médio e alto risco de crédito e que sejam pessoas físicas e jurídicas, correntistas oriundos do BESC.

O objetivo do fundo é acompanhar o CDI utilizando de investimentos em ativos de renda fixa, tendo uma limitação em deter até 50% em ativos privados. Segue abaixo na Figura 3 as condições necessárias para realizar investimento no fundo segundo o Banco do Brasil.

Figura 3 - Condições de Investimento – BANCO DO BRASIL

Investimento inicial mínimo	50000,00
Investimento adicional mínimo	100,00
Resgate mínimo	50,00
Horário para aplicação e resgate	18:00
Valor mínimo para permanência	50,00
Período de carência	Não há
Conversão de cotas	Na aplicação, o número de cotas compradas será calculado de acordo com o valor da cota no fechamento da data da aplicação. No resgate, o número de cotas canceladas será calculado de acordo com o valor da cota no fechamento do dia útil contado da data do pedido de resgate.
Pagamento dos resgates	O prazo para o efetivo pagamento dos resgates é a data do pedido de resgate.

Fonte: BANCO DO BRASIL, 2021.

4.3 Definição dos indicadores a serem comparados.

Os indicadores que serão utilizados nesta análise são IPCA (Índice Nacional de Preços ao Consumidor Amplo), CDI (Certificado de Depósito Interbancário) e Câmbio.

A série histórica do IPCA foi extraída diretamente do IBGE, com dados históricos mensais datados desde 1994, para fundo de estudos utilizaremos apenas os valores de dos meses nos anos de 2018, 2019 e 2020. A série histórica CDI foi fornecida pelo Banco Central com dados de 2015 em diante, porém neste trabalho será analisado apenas o período de 2018 até 2020. A série histórica da variação de câmbio foi fornecida pelo Banco Central já extraídos os dados de 2018 a 2020.

5. RESULTADOS E DISCUSSÕES

Nessa sessão foram discutidos e analisados os resultados obtidos pelos Métodos de Classificação por Naive Bayes, Regressão Linear e a Planilha de Previsão do Excel, com os dois fundos e respectivas base de dados e os três indicadores e suas respectivas bases. Os dados foram organizados com o objetivo verificar a correlação entre determinado fundo e indicadores e entre fundos. Para a análise dos resultados, as classificações foram plotadas mostrando qual o erro com base no histórico da relação.

5.1 Análise de Dados para o fundo CAIXA

Primeiramente na análise do fundo (INV EM COTAS FUNDOS INV CAIXA GERAÇÃO JOVEM RENDA FIXA CRÉDITO PRIVADO LONGO PRAZO), utilizou-se o algoritmo Naive Bayes através da ferramenta WEKA. Foi programado o algoritmo com os índices CDI, Câmbio e IPCA como preditores deste fundo de investimento. Ao rodar o algoritmo, obteve-se de resultado o apresentado na Figura 4.

Figura 4 - Resultados Método de Naive Bayes – Fundo Caixa

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          2           5.5556 %
Incorrectly Classified Instances       34           94.4444 %
Kappa statistic                        -0.0149
Mean absolute error                    0.0833
Root mean squared error                0.2134
Relative absolute error                99.6863 %
Root relative squared error            103.8123 %
Total Number of Instances              36

```

Fonte: O autor (2021)

Pode-se observar que o algoritmo Naive Bayes apresentou um erro de classificação de 94,44% e acerto de 5,56%. Indicando que os preditores escolhidos (CDI, Câmbio e IPCA) não conseguem explicar a variação do fundo de investimento, inviabilizando o uso do algoritmo Naive Bayes.

Uma segunda forma de classificação utilizada, foi o método da Regressão Linear através do Excel para verificação da possibilidade ou não da previsão através de Regressão. Utilizou-se da mesma forma como preditores (CDI, Câmbio e IPCA) para chegar à classificação do fundo de investimento. Os resultados estão demonstrados na Figura 5.

Figura 5 - Resultado Método Regressão Linear - Fundo Caixa vs indicadores

	A	B	C	D	E	F	G	H	I	J
1	RESUMO DOS RESULTADOS									
2										
3	<i>Estadística de regressão</i>									
4	R múltiplo	0,638035909								
5	R-Quadrado	0,407089821								
6	R-quadrado ajustado	0,351504492								
7	Erro padrão	0,002048744								
8	Observações	36								
9										
10	ANOVA									
11		<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>				
12	Regressão	3	9,22203E-05	3,07401E-05	7,323691828	0,000715681				
13	Resíduo	32	0,000134315	4,19735E-06						
14	Total	35	0,000226536							
15										
16		<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>95% inferiores</i>	<i>95% superiores</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>	
17	Interseção	0,0001652	0,001197872	0,137910939	0,891174964	-0,002274786	0,002605185	-0,002274786	0,002605185	
18	Câmbio	-0,023767789	0,007068535	-3,362477493	0,00201515	-0,038165924	-0,009369655	-0,038165924	-0,009369655	
19	CDI	0,698077224	0,246926221	2,82706802	0,008033814	0,19510497	1,201049477	0,19510497	1,201049477	
20	IPCA	0,091961517	0,092547604	0,99366718	0,327839715	-0,096551784	0,280474819	-0,096551784	0,280474819	
21										

Fonte: O autor (2021)

Através do Excel e da ferramenta Análise de Dados obteve-se o resultado da Regressão como sendo 40,70% da variação do fundo de investimento sendo explicado pelos preditores (CDI, Câmbio e IPCA). Ao se analisar o valor P das variáveis pode-se observar que as únicas que apresentaram um resultado abaixo de 5% foram o CDI e Câmbio, indicando que estas variáveis poderiam servir de base para prever o fundo de investimento. Ao apresentar um R-Múltiplo de 63,8%, indica que as variáveis possuem correlação, ou seja, aumentar o valor delas aumentaria também o valor do fundo de investimento. Mas ao apresentar um R-Quadrado baixo (40,70%) nos mostra que apenas em torno de 40% da variação do fundo de investimento pode ser explicada pelos três preditores.

Para corroborar a análise acima, usando a mesma forma de classificação, analisou-se se havia Regressão entre o fundo e CDI, chegando aos valores demonstrados na Figura 6.

Figura 6 - Resultado Método Regressão Linear - Fundo Caixa vs CDI

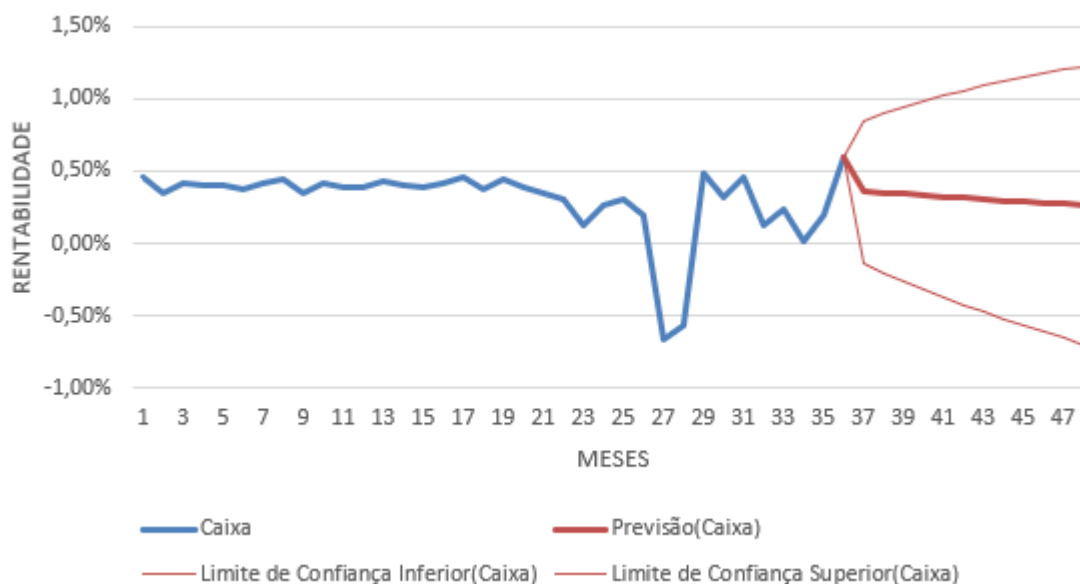
	A	B	C	D	E	F	G	H	I	J
1	RESUMO DOS RESULTADOS									
2										
3	<i>Estadística de regressão</i>									
4	R múltiplo	0,35940055								
5	R-Quadrado	0,129168755								
6	R-quadrado ajustado	0,103556071								
7	Erro padrão	0,002408772								
8	Observações	36								
9										
10	ANOVA									
11		<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>				
12	Regressão	1	2,92613E-05	2,92613E-05	5,043155832	0,031333465				
13	Resíduo	34	0,000197274	5,80218E-06						
14	Total	35	0,000226536							
15										
16		<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>95% inferiores</i>	<i>95% superiores</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>	
17	Interseção	0,00042413	0,001219923	0,347669742	0,730231426	-0,002055051	0,002903312	-0,002055051	0,002903312	
18	CDI	0,63182707	0,281350075	2,245697182	0,031333465	0,060054925	1,203599215	0,060054925	1,203599215	
19										

Fonte: O autor (2021)

Pode-se observar que mesmo utilizando-se da melhor variável explicada pelo valor P na Regressão anterior, nesta nova análise ela apresenta um valor de R-Quadrado baixíssimo, demonstrando que mesmo individualmente os preditores não podem ser usados para prever o resultado do fundo de investimento.

Pode-se perceber que os algoritmos de Aprendizado de Máquina não foram eficientes para prever o fundo de investimento com base nos três preditores escolhidos, chegando a um resultado com erros muito elevado. Desta forma como último teste, fez-se uma análise através do Excel chamado de Planilha de Previsão, que na verdade é uma previsão de séries temporais. Esse recurso utiliza-se de um modelo matemático que faz uma auto regressão dos dados e adiciona-se uma média móvel dos erros com base no histórico dos dados do fundo de investimento, para prever o comportamento futuro, conforme Figura 7.

Figura 7 - Dados reais e previstos pela Planilha de Previsão, aplicado a base de 36 meses – Fundo Caixa



Fonte: O autor (2021)

A linha central vermelha é a previsão para eventos futuros baseados nos próprios dados históricos dos 36 meses do fundo de investimento. As linhas superior e inferior (vermelhas) trata-se do limite superior e inferior dessa previsão.

Na Figura 8, apresenta-se os valores máximos, mínimos e média dos seguintes índices: previsão de rentabilidade futura, o limite de Confiança Inferior e o Limite de Confiança Superior para os próximos 12 meses.

Figura 8 – Dados Previstos Através da Planilha de Previsão – Fundo Caixa

	Previsão(Caixa)	Limite de Confiança Inferior(Caixa)	Limite de Confiança Superior(Caixa)
Máximo	0,36%	-0,13%	1,22%
Mínimo	0,26%	-0,70%	0,85%
Média	0,31%	-0,44%	1,06%

Fonte: O autor (2021)

Apesar do recurso planilha de Previsão não apresentar uma taxa de acerto, ela nos dá um resultado provável baseado no histórico, demonstrando que ainda que incerto, ainda é a melhor forma prever através de séries temporais.

5.2 Análise de Dados para o fundo BB

Primeiramente na análise do fundo (BB BESC RENDA FIXA PRÁTICO CRÉDITO PRIVADO FUNDO DE INVESTIMENTO), utilizou-se o algoritmo Naive Bayes através da ferramenta WEKA. Foi programado o algoritmo com os índices CDI, Câmbio e IPCA como preditores deste fundo de investimento. Ao rodar o algoritmo, obteve-se de resultado o apresentado na Figura 9.

Figura 9 - Resultados Método de Naive Bayes – Fundo BB

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      9           25    %
Incorrectly Classified Instances    27           75    %
Kappa statistic                     0.1811
Mean absolute error                  0.0827
Root mean squared error              0.2096
Relative absolute error              94.9788 %
Root relative squared error          99.8809 %
Total Number of Instances           36

```

Fonte: O autor (2021)

Pode-se observar que o algoritmo Naive Bayes apresentou um erro de classificação de 75% e acerto de 25%. Indicando que os preditores escolhidos (CDI, Câmbio e IPCA) não conseguem explicar a variação do fundo de investimento, inviabilizando o uso do algoritmo Naive Bayes.

Uma segunda forma de classificação empregue, foi o método da Regressão Linear através do Excel para verificação da possibilidade ou não da previsão através de Regressão. Utilizou-se da mesma forma como preditores (CDI, Câmbio e IPCA) para chegar à classificação do fundo de investimento. Os resultados estão demonstrados na Figura 10.

Figura 10 - Resultado Método Regressão Linear - Fundo BB vs indicadores

	A	B	C	D	E	F	G	H	I	J
1	RESUMO DOS RESULTADOS									
2										
3	<i>Estatística de regressão</i>									
4	R múltiplo	0,527366431								
5	R-Quadrado	0,278115353								
6	R-quadrado ajustado	0,210438667								
7	Erro padrão	0,001317536								
8	Observações	36								
9										
10	ANOVA									
11		<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>				
12	Regressão	3	2,14009E-05	7,13363E-06	4,109470643	0,014195782				
13	Resíduo	32	5,55488E-05	1,7359E-06						
14	Total	35	7,69497E-05							
15										
16		<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>95% inferiores</i>	<i>95% superiores</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>	
17	Interseção	0,005927769	0,000770345	7,69495434	9,00894E-09	0,004358628	0,007496911	0,004358628	0,007496911	
18	Câmbio	-0,000763272	0,004545736	-0,167909472	0,867711218	-0,010022633	0,008496089	-0,010022633	0,008496089	
19	CDI	-0,557445714	0,158796893	-3,510432123	0,001353951	-0,8809044	-0,233987028	-0,8809044	-0,233987028	
20	IPCA	-0,054635364	0,059516855	-0,917981368	0,365496213	-0,17586723	0,066596502	-0,17586723	0,066596502	
21										

Fonte: O autor (2021)

Através do Excel e da ferramenta Análise de Dados obteve-se o resultado da Regressão como sendo 27,81% da variação do fundo de investimento sendo explicado pelos preditores (CDI, Câmbio e IPCA). Ao se analisar o valor P das variáveis pode-se observar que a única que apresentou um resultado abaixo de 5% foi o CDI, indicando que esta variável poderia servir de base para prever o fundo de investimento. Ao apresentar um R-Múltiplo de 52,7%, indica que as variáveis possuem correlação, ou seja, ao aumentar o valor delas aumentaria também o valor do fundo de investimento. Mas ao apresentar um R-Quadrado baixo (27,81%) nos mostra que apenas em torno de 28% da variação do fundo de investimento pode ser explicada pelos três preditores.

Para corroborar a análise acima, usando a mesma forma de classificação, analisou-se se havia Regressão entre o fundo e CDI, chegando aos valores demonstrados na Figura 11.

Figura 11 - Resultado Método Regressão Linear – Fundo BB vs CDI

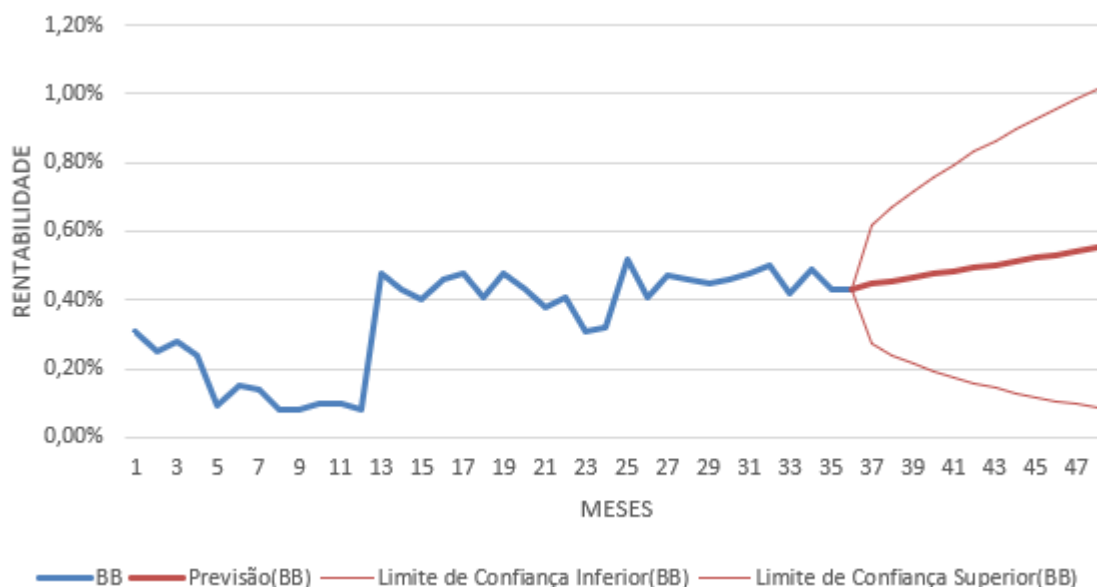
	A	B	C	D	E	F	G	H	I	J
1	RESUMO DOS RESULTADOS									
2										
3	<i>Estadística de regressão</i>									
4	R múltiplo	0,508914437								
5	R-Quadrado	0,258993905								
6	R-quadrado ajustado	0,237199608								
7	Erro padrão	0,001295015								
8	Observações	36								
9										
10	ANOVA									
11		<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>				
12	Regressão	1	1,99295E-05	1,99295E-05	11,88356319	0,001526492				
13	Resíduo	34	5,70202E-05	1,67707E-06						
14	Total	35	7,69497E-05							
15										
16		<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>95% inferiores</i>	<i>95% superiores</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>	
17	Interseção	0,005582206	0,000655861	8,511267818	6,09977E-10	0,004249337	0,006915075	0,004249337	0,006915075	
18	CDI	-0,52143431	0,151260755	-3,447254442	0,001526492	-0,828833149	-0,214035471	-0,828833149	-0,214035471	
19										

Fonte: O autor (2021)

Pode-se observar que mesmo utilizando-se da melhor variável explicada pelo valor P na Regressão anterior, nesta nova análise ela apresenta um valor de R-Quadrado ainda mais baixo, demonstrando que mesmo individualmente os preditores não podem ser usados para prever o resultado do fundo de investimento.

Pode-se perceber que os algoritmos de Aprendizado de Máquina não foram eficientes para prever o fundo de investimento com base nos três preditores escolhidos, chegando a um resultado com erros muito elevado. Desta forma como último teste, fez-se uma análise através do Excel chamado de Planilha de Previsão, que na verdade é uma previsão de séries temporais. Esse recurso utiliza-se de um modelo matemático que faz uma auto regressão dos dados e adiciona-se uma média móvel dos erros com base no histórico dos dados do fundo de investimento, para prever o comportamento futuro, conforme Figura 12.

Figura 12 - Dados reais e previstos pela Planilha de Previsão, aplicado a base de 36 meses – Fundo BB



Fonte: O autor (2021)

A linha central vermelha é a previsão para eventos futuros baseados nos próprios dados históricos dos 36 meses do fundo de investimento. As linhas superior e inferior (vermelhas) trata-se do limite superior e inferior dessa previsão.

Na Figura 13 apresenta-se os valores máximos, mínimos e média dos seguintes índices: previsão de rentabilidade futura, o limite de Confiança Inferior e o Limite de Confiança Superior para os próximos 12 meses.

Figura 13 - Dados Previstos Através da Planilha de Previsão – Fundo BB

	Previsão(Caixa)	Limite de Confiança Inferior(Caixa)	Limite de Confiança Superior(Caixa)
Máximo	0,55%	0,27%	1,02%
Mínimo	0,45%	0,09%	0,62%
Média	0,50%	0,16%	0,84%

Fonte: O autor (2021)

Apesar do recurso planilha de Previsão não apresentar uma taxa de acerto, ela nos dá um resultado provável baseado no histórico, demonstrando que ainda que incerto, ainda é a melhor forma prever através de séries temporais.

6. DISCUSSÕES

Como pode-se observar, os algoritmos de Aprendizado de Máquina não foram eficientes para prever o fundo de investimento através dos preditores escolhidos. Eles não apresentaram um nível de correlação elevado que possa ser considerado para realização de previsão do fundo.

Em ambos os fundos os resultados obtidos foram com erros elevados e taxa de acerto bem baixa, mesmo analisando-se o preditor que possui uma melhor correlação (CDI), nos dois fundos ainda se obteve resultado com grandes erros, ficando distante do necessário para realizar a previsão do fundo de investimento.

Desta forma, na busca por uma corroboração do resultado acima apresentado, utilizou-se uma ferramenta de análise de séries temporais através do Excel, para verificação de uma melhor possibilidade de previsão dos fundos. Conseguiu-se constatar que ainda que distante de um bom resultado, as séries temporais alcançaram resultados melhores de previsão de ambos os fundos, que os algoritmos de Aprendizado de Máquina utilizando os três preditores escolhidos.

Constatou-se como resultado que os algoritmos de Aprendizado de Máquina não foram eficazes no estabelecimento de uma relação entre os fundos de renda fixa e os indicadores econômicos escolhidos, desta forma sendo ineficazes na previsão de retorno desses fundos.

7. CONCLUSÃO

A conclusão deste trabalho nos traz respostas que os fundos não são possíveis de serem previstos através de algoritmos de Aprendizado de Máquinas junto aos preditores selecionados. Verificou-se que não existe uma correlação entre fundos de investimento e grandes indicadores da economia e percebe-se que a análise por Séries Temporais ainda é mais eficaz na previsão da rentabilidade de ambos os fundos.

Desta forma, por meio dessa monografia será possível compreender de forma mais clara como os mecanismos de análise de dados funcionam e entender que muitas vezes os resultados obtidos em sua grande maioria são conclusivos, apesar de não serem o esperado. As análises possuíram erros acima de 70% na classificação das rentabilidades dos fundos, e também apenas 40% do fundo sendo explicado pelos indicadores de mercado em seu melhor resultado, ficando distante do ideal que seriam os resultados que trazem correlação das bases e preditores.

8. REFERÊNCIAS

BERGER, Rafael F. Avaliação da Sensibilidade dos Fundos de Investimento Imobiliários à Variações nas taxas de juros através da análise de componentes principais. Disponível em: <<https://bibliotecadigital.fgv.br/dspace/handle/10438/13492>>. Acesso em 23 abril 2019.

SCHUTT, Isabel Gaio; CALDEIRA, João Frois. Análise de Estilo Dinâmica de Fundos Multimercados: Aplicação para o Mercado Brasileiro. **Análise Econômica**, Porto Alegre, p. 101-129, 1 mar. 2016. Disponível em: <https://www.seer.ufrgs.br/AnaliseEconomica/article/view/45779/36639>. Acesso em: 15 maio 2019.

SARTORI, Anya P. R. **A Utilização de Pesquisas Exploratórias na Área de Marketing**. Disponível em: <<http://eduem.uem.br/laboratorio/ojs/index.php/rimar/article/view/26692/1433>>. Acesso em 15 de maio 2019.

SOUSA, Valmir D.; DRIESSNACK, Martha; COSTA, Isabel A.M. **Revisão dos desenhos de pesquisa relevantes para enfermagem. Parte 1: Desenhos de pesquisa Quantitativa**. Disponível em <http://www.scielo.br/pdf/rlae/v15n3/pt_v15n3a22>. Acesso em 15 de maio 2019.

HIRT, Leandro R. **Análise de Retorno dos Fundos de Renda Fixa Brasileiros Através de Indicadores de Mercado**. Disponível em: <<https://www.lume.ufrgs.br/handle/10183/4922>>. Acesso em 15 de maio 2019.

MENDONÇA, João A. J.; HEITOR, Carlos C.; PEREIRA, Ricardo C. L. **A Escolha de Fundos de Ações e o Investidor Individual**. Disponível em: <<https://www.redalyc.org/pdf/840/84050701004.pdf>>. Acesso em 15 de maio 2019 <http://www.anbima.com.br/pt_br/noticias/renda-fixa-lidera-captacao-dos-fundos-de-investimento-em-janeiro.htm>. Acesso em 29 maio 2019.

MACEDO, M. A. S.; MACEDO, H. D. R. **Avaliação de desempenho de fundos DI no Brasil: uma proposta metodológica**. Revista Economia e Gestão, v. 7, n. 15, 2007. Disponível em: <

<http://periodicos.pucminas.br/index.php/economiaegestao/article/view/9>>. Acesso em 29 maio 2019.

BRITO, N. R.; TACIRO, A. C.; **Características da relação entre taxas selic e cdi e suas implicações.** Disponível em: <http://www.noba.com.br/artigos/Artigo%201-Parte%203- Caracteristicas%20da%20Rela%C3%83%C2%A7_o%20entre%20Taxas%20Selic%20e%20CDI%20e%20suas%20Implica%C3%83%C2%A7_es-05-2002.pdf>. Acesso em 26 de junho 2019.

LEUSIN, L. M. C.; BRITO, R. D.; **Market Timing E Avaliação De Desempenho Dos Fundos Brasileiros.** Disponível em: <www.scielo.br/pdf/rae/v48n2/v48n2a03.pdf>. Acesso em 17 de junho 2019.

MARTINI, M. F. G.; **Renda fixa versus renda variável: uma análise descritiva entre as rentabilidades dos investimentos.** Disponível em <<http://businessstur.com.br/uploads/arquivos/c131f9abdf318c68b9c5f5fb8aa14a9b.pdf>>. Acesso em 01 de julho 2019.

ZINI, A. A. J.; **Taxa de câmbio e política cambial no Brasil.** Disponível em <https://books.google.com.br/books?hl=pt-BR&lr=&id=g6_bbGk3_tQC&oi=fnd&pg=PA11&dq=conceito+de+c%C3%A2mbio&ots=qifFQcFhNP&sig=j_dRgqnpjgMOzi4yRRCDsqNRjG0#v=onepage&q=conceito%20de%20c%C3%A2mbio&f=false>. Acesso em 02 de julho 2019.

BRESSER-PEREIRA, L. C.; **A taxa de câmbio no centro da teoria do desenvolvimento.** Disponível em: <<https://www.revistas.usp.br/eav/article/view/39481/42365>>. Acesso em 16 de julho de 2019.

GARCIA, Márcio G. P. *et al.* TAXA DE JUROS, RISCO CAMBIAL E RISCO BRASIL. **Pesquisa e Planejamento Econômico | PPE**, [S. l.], p. 253-298, 1 ago. 2003. Disponível em: http://repositorio.ipea.gov.br/bitstream/11058/5847/1/PPE_v33_n02_Taxa.pdf. Acesso em: 29 julho 2019.

SILVA, A. L. P.; CARVALHO, E. F.; COSTA, M. O benchmark do mercado de renda fixa brasileiro: análise das propostas de substituição da taxa CDI a partir das

experiências dos países da Zona do Euro. Disponível em: <<http://dspace.mackenzie.br/handle/10899/14594>>. Acesso em 13 de junho de 2019.

ROCHA, M. D. **Análise de desempenho de fundos de renda fixa no Brasil por meio da Análise Envoltória de Dados (DEA)**. 2013. Dissertação (Mestrado em Gestão de Negócios) - Universidade de Brasília, [S. l.], 2013. Disponível em: <<https://repositorio.unb.br/handle/10482/15013>>. Acesso em: 16 nov. 2019.

NETO, J. M. V. **O MERCADO DE RENDA FIXA NO BRASIL conceitos, precificação e risco**. 1. ed. SP - BRASIL: Saint Paul Editora Ltda, 2019. Disponível em: <<https://books.google.com.br/books?hl=pt-BR&lr=&id=LAqUDwAAQBAJ&oi=fnd&pg=PT4&dq=taxa+de+juros+cdi+conceito&ots=L4UcrJbtZz&sig=JulQce2zUjoiH45fOWbY4IHJDSI#v=onepage&q=taxa%20de%20juros&f=false>>. Acesso em: 16 nov. 2019.

FONSECA, R. L. **REPLICAÇÃO DE ÍNDICES DE RENDA FIXA EM CARTEIRAS: CASO DO IMA-GERAL**. 2017. Dissertação (Mestrado em Economia) - Fundação Getulio Vargas, SP - BRASIL, 2017. Disponível em: <<https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/18036/Disserta%20a7%20a3o-%20Renato%20Fonseca.pdf?sequence=1&isAllowed=y>>. Acesso em: 16 nov. 2019.

BOGLE, J. C. **The index mutual fund: 40 years of growth, change and challenge**. *Financial Analysts Journal*, v.72, n.1, p. 9-13, janeiro/fevereiro 2016. Disponível em: <<http://www.theboxistheretoreason.com/wp-content/uploads/2016/03/faj2Ev722En12E5.pdf>>. Acesso em: 16 nov. 2019.

LEMOS, Rodrigo Lopes de. **A FORÇA DOS INDICADORES FUNDAMENTALISTAS: UMA ANÁLISE DA INFLUÊNCIA DOS INDICADORES FUNDAMENTALISTAS PARA O RETORNO DAS AÇÕES QUE COMPÕEM O ÍNDICE BOVESPA**. 2019. Monografia (Bacharelado em Ciências Contábeis) - Universidade Federal do Rio Grande do Norte, Natal RN, 2019. Disponível em: <<https://monografias.ufrn.br/jspui/handle/123456789/9225>>. Acesso em: 17 nov. 2019.

CARRARA, Anieli Fagundes; BARROS, Geraldo Sant'Ana de Camargo. **A Influência do Preço dos Hortifrutícolas no IPCA: uma análise por meio da curva de**

Phillips. **Revista de Economia e Sociologia Rural**, [s. l.], 14 ago. 2016. Disponível em: <http://www.scielo.br/scielo.php?pid=S0103-20032016000400751&script=sci_arttext>. Acesso em: 17 nov. 2019.

BARBOSA, Luiz Matheus. **ANÁLISE DE COMPONENTES PRINCIPAIS NO ESTUDO DO IPCA-15**. 2018. Monografia (Bacharelado em Estatística) - Universidade Federal de Uberlândia, [S. l.], 2018. Disponível em: <<http://repositorio.ufu.br/bitstream/123456789/23957/1/An%c3%a1liseComponentesPrincipais.pdf>>. Acesso em: 17 nov. 2019.

LAMEIRAS, Maria Andréia Parente; CARVALHO, Leonardo Mello de. A INFLAÇÃO DOS ALIMENTOS: UMA ANÁLISE DO DESEMPENHO RECENTE. **Carta de Conjuntura**, [s. l.], 1 jun. 2016. Disponível em: <http://repositorio.ipea.gov.br/bitstream/11058/6917/1/cc31-nt_inflacao.pdf>. Acesso em: 17 nov. 2019

GONÇALVES, Nilson *et al.* Análise da taxa SELIC, PIB E Inflação no Brasil, no período de 2007 a 2016. **Revista de Economia: Empresas e Empreendedores na CPLP**, [S. l.], p. 75-91, 23 mar. 2018. Disponível em: <<https://revistas.ponteditora.org/index.php/e3/article/view/42/37>>. Acesso em: 17 nov. 2019.

RENAULT, Bruno Pontes. **A Relação entre Índices do Mercado Acionário e Taxas de Câmbio: evidências empíricas na América Latina**. 2017. Dissertação (Pós-Graduação em Administração de Empresas) - Pontifícia Universidade Católica do Rio de Janeiro, [S. l.], 2017. Disponível em: <<http://www.lambda.maxwell.ele.puc-rio.br/36259/36259.PDF>>. Acesso em: 18 nov. 2019.

NONNENBERG, Marcelo José Braga *et al.* NOVOS CÁLCULOS DA TAXA EFETIVA REAL DE CÂMBIO PARA O BRASIL. **Carta de Conjuntura**, [S. l.], p. 125-138, 1 set. 2015. Disponível em: <http://repositorio.ipea.gov.br/bitstream/11058/6215/1/nt_cc28_cambio.pdf>. Acesso em: 18 nov. 2019.

SALAZAR, Luiz Filipe Carreiro. **Deteção de Estilos de Aprendizagem em Ambientes Virtuais de Aprendizagem Utilizando Redes Bayesianas**. 2017. Dissertação (Mestrado em Educação) - Universidade Federal dos Vales do

Jequitinhonha e Mucuri, [S. l.], 2017. Disponível em: <http://acervo.ufvjm.edu.br/jspui/bitstream/1/1629/1/luiz_filipe_carreiro_salazar.pdf>. Acesso em: 21 nov. 2019

CAMPOS, Willys S. A.; REIS, Rodrigo Quites; TEIXEIRA, Otávio Noura. Um modelo bayesiano que auxilia na identificação de alunos com dificuldades na aprendizagem de programação de computadores. **VIII Congresso Brasileiro de Informática na Educação**, [s. l.], 2017. Disponível em: <<https://www.br-ie.org/pub/index.php/wie/article/view/8695/6256>>. Acesso em: 21 nov. 2019.

CHAVES, Daniel Bruno Resende; PASCOAL, Livia Maia; BELTRÃO, Beatriz Amorim; LEANDRO, Tânia Alteniza; NUNES, Marília Mendes; SILVA, Viviane Martins da; LOPES, Marcos Venícios de Oliveira. Árvore de classificação para identificação de Padrão respiratório ineficaz em crianças com infecção respiratória aguda. **Revista Eletrônica de Enfermagem**, [S. l.], p. 1-9, 21 dez. 2018. Disponível em: <<https://www.revistas.ufg.br/fen/article/view/45401/33267>>. Acesso em: 23 nov. 2019.

PINTO, Ana Rachel; PARACAMPO, Carla Cristina Paiva; ALBUQUERQUE, Luiz Carlos de. ANÁLISE DO CONTROLE POR REGRAS EM PARTICIPANTES CLASSIFICADOS DE FLEXÍVEIS E DE INFLEXÍVEIS. **Revista Latina de Análisis de Comportamiento**, Veracruz, México, p. 171-194, 1 dez. 2016. Disponível em: <<https://www.redalyc.org/pdf/2745/274520152004.pdf>>. Acesso em: 23 nov. 2019.

SANTOS, Fabricia Damando; BERTCH, Magda; WIVES, Leandro. Classificação de alunos desanimados em um AVEA: uma proposta a partir da mineração de dados educacionais. **Anais do XXVI Simpósio Brasileiro de Informática na Educação (SBIE 2015)**, [S. l.], p. 1052-1061, 1 jun. 2015. Disponível em: <<https://br-ie.org/pub/index.php/sbie/article/view/5418>>. Acesso em: 24 nov. 2019.

RODRIGUES, Rodrigo Lins; MEDEIROS, Francisco P. A. de; GOMES, Alex Sandro. Modelo de Regressão Linear aplicado à previsão de desempenho de estudantes em ambiente de aprendizagem. **II Congresso Brasileiro de Informática na Educação (CBIE 2013)**, [s. l.], p. 607-616, 2013. Disponível em: <<https://br-ie.org/pub/index.php/sbie/article/view/2539/2197>>. Acesso em: 24 nov. 2019.

BORGES, Vinicius Ruela Pereira. **Comparação entre as Técnicas de Agrupamento K-Means e Fuzzy C-Means para Segmentação de Imagens Coloridas**. 2010. Artigo (Mestrado em Computação) – Universidade Federal de Uberlândia (UFU). Disponível em: <https://www.enacomp.com.br/2010/anais/artigos/completos/enacomp2010_42.pdf>. Acesso em: 24 nov. 2019.

VARGAS, Rogério R. de; BEDREGAL, Benjamim Rene Callejas; PALMEIRA, Eduardo Silva. **A Comparison between K-Means, FCM and ckMeans Algorithms**. 2011. Artigo (Doutorado em métodos de Modelagem) - Workshop-School on Theoretical Computer Science, [S. l.], 2011. Disponível em: <<https://ieeexplore.ieee.org/xpl/conhome/6114270/proceeding>>. Acesso em: 24 nov. 2019.

ZEVIANI, Walmes M.; JÚNIOR, Paulo J. R.; BONAT, Wagner H, 2013. Disponível em: <<https://www.ime.unicamp.br/~cnaber/cursomodelosnaolinearesR.pdf>>. Acesso em: 25 nov. 2019.

PANAGGIO, Bruna Z.; MARCACINI, Ricardo M.; REZENDE, Solange O.. Torch-ETS: **análise exploratória de tópicos emergentes com apoio de agrupamento hierárquico de textos**. In: WORKSHOP DE FERRAMENTAS E APLICAÇÕES - SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 2011, Florianópolis. Anais Estendidos do XVII Simpósio Brasileiro de Sistemas Multimídia e Web. Porto Alegre: Sociedade Brasileira de Computação, oct. 2011. p. 91-93. ISSN 2596-1683. Disponível em: <https://sol.sbc.org.br/index.php/webmedia_estendido/article/view/4999>. Acesso em: 25 nov. 2019.

MURRUGARRA-LLERENA, Nils; ALVA-MANCHEGO, Fernando; REZENDE, Solange Oliveira. **Comparação de Grades Curriculares de Cursos de Computação Baseada em Agrupamento Hierárquico de Textos**. 2011. Artigo (Instituto de Ciências Matemáticas e de Computação) - Universidade de São Paulo, [S. l.], 2011. Disponível em: <http://zeus.dimap.ufrn.br/csbc2011/anais/eventos/contents/WEI/Wei_Secao_4_Artigo_3_Murrugarra-Llerena.pdf>. Acesso em: 25 nov. 2019.

CARDOZO, Tuane et al. **Análise do Perfil de Investidores Brasileiros**. IX CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO, Ponta Grossa - PR, 4 dez. 2019. Disponível em: <http://aprepro.org.br/conbrepro/2019/anais/arquivos/09292019_140900_5d90e7fc62875.pdf>. Acesso em: 7 março 2021.

ÁRIAS, JOÃO CARLOS GARCIA. **TESTE DE APLICAÇÕES BASEADO EM ANÁLISE DE INSTÂNCIAS DE DADOS ALTERNATIVAS**. Orientador: Prof.^a Dr.^a Silvia Regina Vergilio. 2011. 94 p. Trabalho de Conclusão de Curso (Mestrado em Informática) - UNIVERSIDADE FEDERAL DO PARANÁ – UFPR, [S. l.], 2011. Disponível em: <<https://www.acervodigital.ufpr.br/handle/1884/25833>>. Acesso em: 19 abril 2021.

LOPES, Marcela Galvão *et al.* ANÁLISE DOS INDICADORES DE INADIMPLÊNCIA NAS LINHAS DE CRÉDITO PARA PESSOA FÍSICA: UM ESTUDO UTILIZANDO MODELO DE REGRESSÃO LOGÍSTICA. **Estudos do CEPE**, [S. l.], p. 75-90, 30 nov. 2017. Disponível em: <<https://online.unisc.br/seer/index.php/cepe/article/view/11099/6866>>. Acesso em: 14 maio 2021.

GOUVÊA, Maria Aparecida *et al.* APLICAÇÃO DE REGRESSÃO LOGÍSTICA E ALGORITMOS GENÉTICOS NA ANÁLISE DE RISCO DE CRÉDITO. **Revista Universo Contábil**, [S. l.], p. 84-102, 1 abr. 2012. Disponível em: <<https://www.redalyc.org/pdf/1170/117023653005.pdf>>. Acesso em: 14 maio 2021.

CAMARGOS, M. A.; CAMARGOS, M. C. S.; ARAÚJO, E. A. A inadimplência em um programa de crédito de uma instituição financeira pública de Minas Gerais: uma análise utilizando regressão logística. **REGE Revista de Gestão**, v. 19, n. 3, p. 467-486, jul./set. 2012. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1809227616303204>>. Acesso em: 15 de maio 2021.

LORENA, Ana Carolina. **Investigação de estratégias para a geração de máquinas de vetores de suporte multiclases**. 2006. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - USP São Carlos, [S. l.], 2006. Disponível em: <<https://www.teses.usp.br/teses/disponiveis/55/55134/tde-26052006-111406/en.php>>. Acesso em: 16 maio 2021.

RAUBER, Thomas Walter. **Redes Neurais Artificiais**. 2005. Tese (Monografia) - UFES - Universidade Federal do Espírito Santo, [S. l.], 2005. Disponível em: <https://www.researchgate.net/profile/Thomas-Rauber-2/publication/228686464_Redес_neurais_artificiais/links/02e7e521381602f2bd000000/Redes-neurais-artificiais.pdf>. Acesso em: 16 maio 2021.

SILVA, Leandro A. *et al.* Prática de Mineração de Dados no Exame Nacional do Ensino Médio. **3º Congresso Brasileiro de Informática na Educação**, [S. l.], p. 651-660, 1 jun. 2014. Disponível em: <[mas-Rauber-2/publication/228686464_Redес_neurais_artificiais/links/02e7e521381602f2bd000000/Redes-neurais-artificiais.pdf](https://www.researchgate.net/profile/Thomas-Rauber-2/publication/228686464_Redес_neurais_artificiais/links/02e7e521381602f2bd000000/Redes-neurais-artificiais.pdf)>. Acesso em: 16 maio 2021.

DIAS, Bruna Baldini. **Os métodos bayesianos de aprendizado de máquina pelos algoritmos naïve bayes e redes de crença na predição de período chuvoso na cidade de Blumenau**. 2018. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade do Extremo Sul Catarinense, [S. l.], 2018. Disponível em: <<http://repositorio.unesc.net/handle/1/8150>>. Acesso em: 16 maio 2021.

MARIANO, Marcos Alves. **Comparação de Algoritmos Paralelos para a Extração de Regras de Associação no Modelo de Memória Distribuída**. 2011. Dissertação (Mestrado em Ciência da Computação) - UFMS - Universidade Federal de Mato Grosso do Sul, [S. l.], 2011. Disponível em: <<https://repositorio.ufms.br/handle/123456789/1674>>. Acesso em: 17 maio 2021.

CAIXA ECONOMICA FEDERAL. *In*: **CAIXA FIC GERACAO JOVEM RF CRED PRIV LP 10.577.485/0001-34**. [S. l.], 1 maio 2021. Disponível em: <<https://www.caixa.gov.br/fundos-investimento/renda-fixa/fic-geracao-jovem-rf-credito-privado-longo-prazo/Paginas/default.aspx>>. Acesso em: 27 jun. 2021.

ANBIMA - ASSOCIAÇÃO BRASILEIRA DAS ENTIDADES DOS MERCADOS FINANCEIRO E DE CAPITAIS. *In*: **CAIXA FIC GERACAO JOVEM RF CRED PRIV LP**. [S. l.], 27 jun. 2021. Disponível em: <<https://data.anbima.com.br/fundos/302181#dadosCadastrais>>. Acesso em: 27 jun. 2021.

MAIS RETORNO. **FIC FI CAIXA GERAÇÃO JOVEM RF CP LP**. [S. l.], 27 jun. 2021. Disponível em: <<https://maisretorno.com/fundo/fic-fi-caixa-geracao-jovem-rf-cp-lp>>. Acesso em: 27 jun. 2021.

BANCO DO BRASIL. **DEMONSTRAÇÃO DE DESEMPENHO DO Besc RF Pratico Cred Privado**. [S. l.], 2020. Disponível em: <<https://www.bb.com.br/docs/pub/siteEsp/dtvm/dwn/sem00073041.pdf>>. Acesso em: 27 jun. 2021.

MINISTÉRIO DA FAZENDA. BANCO CENTRAL. **Relatório de Economia Bancária 2020**. [S. l.: s. n.], 2020. Disponível em: <<https://static.poder360.com.br/2021/06/economia-bancaria-2020-7jun2021.pdf>>. Acesso em: 27 jun. 2021.

FERRARI, Hamilton. **5 maiores bancos detinham 81,8% do mercado de crédito em 2020, diz BC**. [S. l.], 7 jun. 2021. Disponível em: <<https://www.poder360.com.br/economia/5-maiores-bancos-detinham-818-do-mercado-de-credito-em-2020-diz-bc/>>. Acesso em: 27 jun. 2021.

MAIS RETORNO. **BB BESC RF PRÁTICO CP FI**. [S. l.], 27 jun. 2021. Disponível em: <<https://maisretorno.com/fundo/bb-besc-rf-pratico-cp-fi>>. Acesso em: 27 jun. 2021.

BANCO DO BRASIL. **LÂMINA DE INFORMAÇÕES ESSENCIAIS SOBRE O BB Besc Renda Fixa Pratico Crédito Privado**. [S. l.], 1 maio 2021. Disponível em: <<https://www.bb.com.br/docs/pub/siteEsp/dtvm/dwn/lam00073041.pdf>>. Acesso em: 27 jun. 2021.

In: INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **IPCA - Índice Nacional de Preços ao Consumidor Amplo**. [S. l.], 30 dez. 2020. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo.html?=&t=series-historicas>. Acesso em: 28 junho 2021.

MINISTÉRIO DA FAZENDA. BANCO CENTRAL. **Relatório de Economia Bancária 2020**. [S. l.: s. n.], 2020. Disponível em: <<https://blog.nubank.com.br/cdi-2021/>>. Acesso em: 29 junho 2021.

DEBIT. **CDI - HISTÓRICO COMPLETO**. [S. /], 1 jul. 2021. Disponível em: <<https://www.debit.com.br/tabelas/tabela-completa.php?indice=cdi>>. Acesso em: 29 junho 2021.

FUSION MEDIA LIMITED. **USD/BRL Dados Históricos**. [S. /], 4 jul. 2021. Disponível em: <<https://br.investing.com/currencies/usd-brl-historical-data>>. Acesso em: 1 julho 2021.

UNIVERSIDADE DE WAIKATO. **Weka 3: Machine Learning Software in Java**. [S. /], 2021. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>. Acesso em: 1 julho 2021.

AMY, Antonio Sergio Azevedo. **COMO CLASSIFICAR AS PESQUISAS?. ACADEMIA- Accelerating the world's research.**, [S. /], p. 1-15, 19 ago. 2021. Disponível em: <https://d1wqtxts1xzle7.cloudfront.net/38881088/como_classificar_pesquisas-with-cover-page-v2.pdf?Expires=1629407736&Signature=GsfTl0TuXkRiK6U9sb3tSabPY9IBYQUiR6k5ZHjiUEEtXalzv6Xho6pTz9Qf5YawHA9X22JzXrWiwuUMcj-4-wjXbFwqX3rQx2OPZGK7XMpk-5itwwk8Urn139WYJzfrt4LW5yizaRwFlqLMA4LnIwMw~CP0IFG~I1Ela1E6UQrhwbDqJ5hLpXH5attUx8uScs4e31NKn6G5sSu8zeVwuNjbl2wx3QB8YINyCH2OSXSs4QNA t67Qy8t4j~qe10i-h6AhnxFungo4m~3flv6JBw2NcgzK8EFfswHCVkVIqr8XMsKwraEVxSk-kcl4DfdXsbOlwgwIDIWi4jvD~hyV4w__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA>. Acesso em: 15 ago. 2021.