

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**

**DOUGLAS TEIXEIRA COBIANK**

**MODELO DE SISTEMA DE RECOMENDAÇÃO PARA RECUPERAÇÃO DE  
CRÉDITO COM FOCO NA EQUIPE DE COBRANÇA**

**PONTA GROSSA**

**2022**

**DOUGLAS TEIXEIRA COBIANK**

**MODELO DE SISTEMA DE RECOMENDAÇÃO PARA RECUPERAÇÃO DE  
CRÉDITO COM FOCO NA EQUIPE DE COBRANÇA**

**A recommendation system model for credit recovery focused on the collection team**

Trabalho de conclusão de curso de graduação apresentado como requisito para obtenção do título de Bacharel em Ciência da Computação da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador(a): Prof. Luiz Rafael Schmitke

**PONTA GROSSA**

**2022**



[4.0 Internacional](https://creativecommons.org/licenses/by-sa/4.0/)

Esta licença permite remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

**DOUGLAS TEIXEIRA COBIANK**

**MODELO DE SISTEMA DE RECOMENDAÇÃO PARA RECUPERAÇÃO DE  
CRÉDITO COM FOCO NA EQUIPE DE COBRANÇA**

Trabalho de conclusão de curso de graduação apresentada como requisito para obtenção do título de Bacharel em Ciência da Computação da Universidade Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 23 /maio/2022

---

Luiz Rafael Schmitke  
Mestre

Universidade Tecnológica Federal do Paraná – Campus Ponta Grossa

---

André Pinz Borges  
Doutor

Universidade Tecnológica Federal do Paraná – Campus Ponta Grossa

---

Helyane Bronoski Borges  
Doutora

Universidade Tecnológica Federal do Paraná – Campus Ponta Grossa

**PONTA GROSSA**

**2022**

## AGRADECIMENTOS

Este trabalho é dedicado a todos aqueles que fizeram parte dessa jornada, mas especialmente ao meus pais e irmãs, que mesmo longe sempre acreditaram em mim e me deram toda força e suporte necessário para seguir em frente.

Toda jornada tem um fim, mas este espero que seja o começo de uma caminhada de muito sucesso. Mas antes de finalizar um ciclo e iniciar outro, gostaria de agradecer todas as pessoas que fizeram parte disso.

Aos meus amigos que diante de todas as dificuldades que enfrentamos na faculdade e na vida, sempre estiveram comigo e foram minha segunda família em Ponta Grossa. Aos meus professores que tive a oportunidade de aprender, compartilhar ideias e admirar cada um por tamanha sabedoria.

Mas principalmente aos meus familiares, a minha mãe Marilene Teixeira Cobiank que sempre rezou e me deu todo o amor e educação necessária para ser o homem que sou hoje, ao meu pai Ademir Cobiank que mesmo com toda a dificuldade pode me auxiliar financeiramente e amorosamente durante todo o período em que estive na universidade, as minhas irmãs Juliane e Joselaine que vieram me visitar, foram minhas confidentes, parceiras e melhores amigas.

Gostaria também de agradecer ao meu orientador Prof. Luiz Rafael Schmitke que mesmo sem me conhecer muito bem, aceitou esse desafio e pode me orientar da melhor forma possível. Vocês foram muito importantes para essa realização.

Apenas o meu muito obrigado.

## RESUMO

Com o aumento do endividamento ano a ano, o Brasil não vive um bom momento econômico, nem mesmo por motivos políticos, econômicos, de saúde ou outros. Apesar disso, o povo brasileiro sempre busca estar resolvendo seus problemas financeiros. No entanto, quando o pagamento das dívidas não é possível, os serviços de proteção ao crédito passam a fazer parte do processo e as dívidas são protestadas judicialmente. Levando em consideração que o avanço da tecnologia da informação e a inteligência de dados possibilitaram o desenvolvimento de ferramentas que auxiliam as pessoas em seu dia a dia, este trabalho propõe a construção de um modelo de recomendação de ações baseado em classificação multirrótulo para facilitar o processo de cobrança de dívidas. Tal modelo preconiza ações que as operadoras podem realizar no momento da cobrança, e que podem resultar em maior assertividade no contato com os devedores, ampliando a possibilidade de quitação das dívidas. Neste trabalho foram criados dois modelos, com mais rótulos e menos rótulos, para realizar a classificação multirrótulo, sendo o primeiro, com menos rótulos, estatisticamente melhor quando comparado ao segundo. A acurácia no primeiro modelo foi em média de 71%, o que indica que, caso a equipe de cobrança tome as ações recomendadas, existe 71% de probabilidade de que a dívida seja quitada. Dessa forma, o modelo de recomendação baseado em classificação multirrótulo alcançou melhores resultados que a taxa normal de recuperação de dívidas, 47%, o que demonstra que pode ser uma alternativa para equipes de cobrança no momento de recuperação de crédito.

Palavras-chave: sistemas de recomendação; multirrótulo; dívidas; *collection score*.

## ABSTRACT

With the increase in indebtedness year after year, Brazil is not experiencing a good economic moment, not even for political, economic, health or other reasons. Despite this, the Brazilian people always seek to solve their financial problems. However, when debt payment is not possible, credit protection services become part of the process and debts are protested in court. Taking into account that the advancement of information technology and data intelligence enabled the development of tools that help people in their daily lives, this work proposes the construction of a model of recommendation of actions based on multi-label classification to facilitate the process of debt collection. This model advocates actions that operators can carry out at the time of collection, which can result in greater assertiveness in contacting debtors, expanding the possibility of paying off debts. In this work, two models were created, with more labels and fewer labels, to perform the multi-label classification, the first, with fewer labels, being statistically better when compared to the second. The accuracy in the first model was on average 71%, which indicates that, if the collection team takes the recommended actions, there is a 71% probability that the debt will be paid off. In this way, the recommendation model based on multi-label classification achieved better results than the normal debt recovery rate, 56%, which demonstrates that it can be an alternative for collection teams at the time of credit recovery.

Keywords: recommendation system; multilabel; debts; collection score.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Ciclo de Crédito.....	12
Figura 2 - Exemplo de decomposição Binaria do classificador BR .....	18
Figura 3 - Estrutura do <i>Classifier Chain</i> .....	19
Figura 4 - Estrutura do Label Powerset.....	20
Figura 5 - Modelo de Filtragem Baseada em Conteúdo .....	23
Figura 6 - Exemplo de Recomendação baseada filtragem colaborativa .....	24
Figura 7 - Equação de Precisão .....	26
Figura 8 - Definição do cálculo do recall .....	26
Figura 9 - Cálculo do F-measure.....	26
Figura 10 - Modelo Relacional da base de Devedores.....	28
Figura 11 - Modelo lógico da base de devedores .....	29
Figura 12 - Script para seleção dos dados .....	30
Figura 13 - Exemplo de arquivo com extensão ARFF .....	35
Figura 14 - Medida de Acurácia para o Conjunto A .....	39
Figura 15 - Métrica de Acurácia para o Conjunto B .....	39
Figura 16 - Medida de Micro Precisão para o Conjunto A .....	40
Figura 17 - Medida de Micro Precisão para o Conjunto B .....	40
Figura 18 - Medida de Micro <i>Recall</i> para o Conjunto A .....	41
Figura 19- Medida de Micro Recall para o Conjunto B .....	41
Figura 20 - Medida de Macro Precisão para o Conjunto A .....	42
Figura 21- Medida de Macro Precisão para o Conjunto B.....	42
Figura 22 - Medida de Macro Recall para o Conjunto A .....	43
Figura 23 - Medida de Macro Recall para o Conjunto B .....	43
Figura 24 - Medida de F1( <i>Micro Averaged</i> ) do Conjunto A .....	44
Figura 25- Medida de F1(Micro Averaged) do Conjunto B .....	44
Figura 26 - Medida de F1 ( <i>Macro Avarage by Exemple</i> ) do Conjunto A.....	45
Figura 27 - Medida de F1 (Macro Avarage by Exemple) do Conjunto B .....	45
Figura 28 - Medida de F1 ( <i>Macro Avarage by Label</i> ) do Conjunto A.....	46
Figura 29 - Medida de F1 (Macro Avarage by Label) do Conjunto B .....	46

## LISTA DE QUADROS E TABELAS

<b>Tabela 1 - Métricas de avaliação para o algoritmo BR/Conjunto A.....</b>	<b>36</b>
<b>Tabela 2 - Métricas de avaliação para o algoritmo BR/Conjunto B.....</b>	<b>36</b>
<b>Tabela 3 - Métricas de avaliação para o algoritmo CC/Conjunto A .....</b>	<b>37</b>
<b>Tabela 4 - Métricas de avaliação para o algoritmo CC/Conjunto B.....</b>	<b>37</b>
<b>Tabela 5 - Métricas de avaliação para o algoritmo LC/Conjunto A.....</b>	<b>38</b>
<b>Tabela 6 - Métricas de avaliação para o algoritmo LC/Conjunto B.....</b>	<b>38</b>
<b>Quadro 1 - Dados da Tabela Forma de Contato .....</b>	<b>30</b>
<b>Quadro 2 - Dados da Tabela Forma de Pagamento .....</b>	<b>31</b>
<b>Quadro 3 - Visualização contendo 10 dados gerados para a base de devedores no SQL Server .....</b>	<b>32</b>



## SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO .....</b>	<b>10</b>
<b>1.1</b>	<b>Objetivos.....</b>	<b>11</b>
<b>2.</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>12</b>
<b>1.2</b>	<b>Ciclo do Crédito .....</b>	<b>12</b>
1.2.1	Collection Score .....	14
<b>1.3</b>	<b>Aprendizagem de Máquina .....</b>	<b>16</b>
1.3.1	Aprendizado Supervisionado.....	16
1.3.2	Classificação Multirrótulo .....	17
<b>1.4</b>	<b>Sistemas de Recomendação.....</b>	<b>21</b>
1.4.1	Filtragem Baseada em Conteúdo.....	22
1.4.2	Filtragem Colaborativa .....	23
1.4.3	Filtragem Híbrida .....	24
<b>1.5</b>	<b>Métricas para Avaliação .....</b>	<b>26</b>
<b>1.6</b>	<b>Trabalhos Similares.....</b>	<b>27</b>
<b>3.</b>	<b>DESENVOLVIMENTO.....</b>	<b>28</b>
<b>4.</b>	<b>RESULTADOS E DISCUSSÕES .....</b>	<b>36</b>
<b>5.</b>	<b>CONCLUSÃO.....</b>	<b>48</b>
	<b>REFERÊNCIAS.....</b>	<b>49</b>

## 1. INTRODUÇÃO

Hoje o Brasil conta com mais de 77,7% de sua população endividada, segundo pesquisa divulgada pela Confederação Nacional do Comércio (CNC) neste ano (O GLOBO, 2022). Esse número vem aumentando cada vez mais por conta de diversos fatores, mas o maior causador desse aumento no último ano foi a pandemia da covid-19. Com mais de 10,5% de desemprego (G1, 2022) o Brasil passa por uma crise econômica sem precedentes, pois comparado com o mesmo período do ano de 2020 o desemprego aumentou em cerca de 3,4% segundo o Instituto Brasileiro de Geografia e Estatística (IBGE) que estima que é a maior taxa desde o início da série histórica (UOL, 2021).

Mas, apesar do desemprego e nível de incerteza no país crescer, o brasileiro busca quitar suas dívidas em médio e longo prazo. Pesquisa recente mostra que cerca de 48% dos empréstimos procurados tem por finalidade o pagamento de dívidas, e muitas vezes a tecnologia pode auxiliar nessa organização para realização dos pagamentos (MACIEL, 2021).

Percebe-se que cada vez mais nossa vida gira em torno de algoritmos e inteligência artificial. Toda a capacidade computacional envolvida nas nossas vidas é percebida de diversas formas, desde qual a roupa da moda, qual a música mais tocada, a comida mais pedida em restaurantes, quem são nossos amigos mais próximos etc. A tecnologia possibilitou que os dados pudessem dizer cada vez mais sobre nós, e como isso afeta nossas vidas.

De acordo com Rich e Knight (1991) o aprendizado de máquina faz com que sejam orientadas informações para o computador de modo que ele possa melhorar gradualmente seu desempenho, isso quer dizer que utilizando de algoritmos focados em dados, estes podem apontar e decidir os melhores fluxos para o nosso problema.

Possíveis aplicações da Aprendizagem de Máquina (AM) são, por exemplo, no mercado de ações, no mercado de empréstimos, no mercado varejista, mas um dos principais pontos da utilização da aprendizagem é conhecer o cliente final, ou *Know Your Customer* (KYC), este processo é um diferencial para o serviço ou empresa apresentar da melhor maneira soluções ou produtos (FORTI, 2018). Essas informações podem dizer quais são os hábitos de pagamento, determinar riscos para as empresas de crédito, se o crédito pode ser recuperado ou não.

Já os sistemas de recomendação, que inicialmente eram chamados de sistemas de filtragem coletiva (MEDEIROS, 2013), conseguem trazer informações relevantes a partir de um mecanismo de notas, que serve como uma propaganda boca a boca. Empresas como a

Amazon, Youtube, Instagram, Netflix e outras, fazem uso desse mecanismo para conseguirem entregar para o usuário final o melhor conteúdo que ele possa estar procurando.

Tendo em vista o aumento do endividamento, o uso de técnicas de AM e sistemas de recomendação podem ser um diferencial competitivo para empresas que hoje trabalham com recuperação de crédito, pois além de poder indicar quais são os clientes que têm maior probabilidade de acertar suas dívidas, também pode ser usada para destacar quais dívidas podem ser mais bem negociadas no dia, e até mesmo fazer previsões sobre como será o endividamento daqui alguns anos.

Dessa forma, uma abordagem possível e que está presente na AM e que pode auxiliar na recomendação de ações para recuperação de crédito é a classificações multirrótulo. Esta visa associar diversos rótulos a uma instância usando suas características, permitindo assim relacionar, no caso deste trabalho, cada rótulo com uma ação a ser recomendada para atingir a recuperação de crédito.

Assim, este trabalho tem como objetivo a criação de um modelo de sistema de recomendação para recuperação de crédito do ponto de vista da equipe de cobrança.

## **1.1 Objetivos**

O objetivo geral do trabalho é desenvolver um modelo de sistema de recomendação de ações para recuperação de crédito utilizando a classificação multirrótulo e os objetivos específicos são:

- Criar uma base de dados sintética de dívidas
- Criar um conjunto de dados multirrotulado de dívidas
- Avaliar o modelo de sistema de recomendação através de métricas para sistemas de recomendação

## 2. FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta a fundamentação teórica necessária para o entendimento dos temas que envolvem o trabalho, como, o ciclo de crédito, classificação multirrótulo e sistemas de recomendação.

### 1.2 Ciclo do Crédito

Consumir sempre foi algo inerente ao ser humano, desde a criação de um sistema de trocas. Este processo ficou cada vez mais complexo com o passar dos anos, principalmente por conta das tecnologias e possibilidades de consumo (MACHADO e NIKOLAY, 2017). Com o surgimento da moeda, a necessidade de definir o tempo para quitação de uma dívida também foi criada, isso foi chamado de crédito, e é o meio de antecipar o poder de compra. O crédito foi criado com base na confiança que o credor, ou seja, quem empresta o dinheiro que o devedor quem pega esse dinheiro emprestado que a dívida gerada será paga, ou seja, um o acordo será cumprido. A existência do crédito é de extrema importância para a estrutura capitalista, pois um indivíduo com dinheiro pode fornecer crédito para um com menos poder, assim cobrando juros e sustentando o ciclo de crédito (NEVES e DEUS, 2017). As instituições credoras realizam diversas análises para identificar os melhores clientes e assim poder oferecer crédito na esperança de que esses valores sejam retornados, sempre existe o risco de inadimplência, mas para isso foi criado o Ciclo de Crédito.

O ciclo de crédito conforme mostrado na Figura 1 é um processo que existe em diferentes mercados, em todo o negócio que deseja oferecer crédito aos seus clientes, mas tendo segurança na hora de conceder tal benefício.

Figura 1 – Ciclo de Crédito



**Fonte: Zuñiga (2019)**

Entre os anos de 1980 e 2003 o mercado de crédito brasileiro passou por um momento de estagnação, mas a partir de 2003 o mercado percebeu grande mudança em como as pessoas lidavam com o crédito e passou a investir cada vez mais nesse setor, contudo após o ano de 2014 foi percebido um grande aumento no endividamento e não pagamento de contas por parte de clientes (PAULA e JUNIOR, 2020). Para minimizar o impacto que existe no não recebimento de valor, as empresas empregam o que chamamos de ciclo de crédito, e este ciclo funciona da seguinte forma (ZUÑIGA, 2019).

- **Prospecção do Cliente:** É levantado um perfil de cliente ideal, ou seja, um cliente que vai consumir um produto ou serviço, este ideal vai levar em conta diversos fatores, mas pode incluir profissão, estado e cidade onde o cliente mora entre outras informações que vai fazer com que seja avaliado o cliente ideal e ele seja confiável para entrar neste ciclo.
- **Concessão de Crédito:** Nesta fase vai ser verificado o risco da concessão de crédito para o cliente, esse risco é calculado a partir da prospecção, aqui será feita a precificação e o limite de crédito concedido.
- **Gestão de Risco:** A gestão de risco vai cuidar de acompanhar o cliente, onde podem ser feitas negociações, oferecidas novas oportunidades e assim um relacionamento entre a empresa e o cliente vai sendo criado para que no final o crédito oferecido seja quitado.
- **Cobrança ou Recuperação:** Mesmo com todas as análises realizadas, sempre existe o risco de existir um cliente inadimplente, então será realizada a etapa de recuperação, nesta etapa a empresa vai oferecer as melhores condições e até descontos para que a dívida seja quitada.

Neste ciclo uma das etapas mais importantes é a recuperação de crédito, pois apesar de haver toda a análise para que os melhores clientes sejam selecionados, sempre pode haver algum fator que impeça o recebimento desse débito. Segundo a Serasa Experian no ano de 2021, mais de 46,5% dos débitos de empresas foram recuperados após um prazo de até 60 dias de negativação (SERASA, 2022).

Em levantamento feito pela Serasa 28,6% dessas dívidas são provenientes dos cartões de crédito, seguidos por dívidas domésticas como água, luz e gás que representam 23,2% e por

último comércio varejistas com 12,5% (SERASA, 2022). Ainda sobre este levantamento é mostrado que o Estado do Amazonas é o estado com mais pessoas inadimplentes tendo cerca de 52,40% da sua população endividada, seguido pelos estados do Distrito Federal com 49%, Rio de Janeiro com 48,19%, Mato Grosso com 47,94% e Amapá com 47,64% (SERASA, 2022). Entre os maiores devedores estão o público feminino que representa 50,2%, mas que também foi o que mais renegociou suas dívidas, chegando em 54,27% das renegociações.

A faixa etária com maior número de inadimplente é a faixa entre 26 e 40 anos tendo cerca de 35,2% de todos os inadimplentes e a faixa que mais negociou foi a faixa de 31 a 40 anos com 31,18% das negociações (SERASA, 2022). O objetivo do trabalho é a utilização da etapa 4, que é a etapa de recuperação de crédito, e dentro dessa etapa existe um mecanismo chamado *collection score*.

### 1.2.1 Collection Score

O *Collection Score* é um sistema que tem como objetivo classificar grupos de clientes inadimplentes entre os com maior e menor chance de quitar suas dívidas, este modelo tem como objetivo a diminuição de custos e assertividade na recuperação financeira. Este processo é aplicado no final do ciclo de crédito, pois como o cliente já está inadimplente, então resta apenas a recuperação de crédito e isso faz com que relacionamento com o cliente se mantenha e se torne um meio de prevenção a perdas.

Pode-se classificar o *Collection Score* em subclasses, onde cada classe representa faixas onde as cobranças são feitas. Estas faixas são Cobrança Preventiva, Modelo de Autocura, *Collection Score*, *Write-OFF*, melhor momento de contato e renegociação.

- Cobrança Preventiva: É utilizada para identificar possíveis clientes que possam entrar em atraso, nisso a empresa pode entrar em contato com os clientes de maior risco com lembretes e pode oferecer outras soluções para não correr o risco de não resolver.
- Modelo de Autocura: É utilizado para identificar pagamentos realizados por clientes com atraso de até 15 dias, o objetivo é evitar a cobrança de clientes que sempre pagam suas contas em dia, mas que por algum imprevisto deixaram alguma dívida de poucos dias para trás, mas que com lembretes enviados pela cobrança preventiva realizaram o pagamento.
- *Collection Score*: Neste modelo são identificadas as chances que um cliente tem de quitar suas dívidas, normalmente é feito entre o 15º dia e o 180º dia de atraso.

- Os modelos de *Write-OFF*: são gerados a partir de clientes que já estão inadimplentes, para recuperar esse crédito as empresas buscam disponibilizar descontos para que a dívida seja quitada.
- Modelo de melhor momento de contato: Como o próprio nome diz, vai ser identificado o melhor meio e momento para entrar em contato, este procedimento pode ser feito durante todo o ciclo da cobrança, e o contato pode ser feito por e-mail, telefone entre outros.
- Modelo de renegociação: É onde são geradas estimativas do cliente realizar a renegociação ou então quebrar o acordo de renegociação pré-estabelecido. e cessão de carteira, este modelo vai dar preços para a carteira e suas dívidas, com isso é possível dizer qual será o percentual recuperado e o valor de venda da carteira (FORTI, 2018).

Para este trabalho será utilizado o *Collection Score* pois o trabalho tem como objetivo a última fase de crédito onde o *collection score* está inserido e o método funciona da seguinte maneira:

- Os clientes são colocados em subgrupos onde indivíduos com prazo de atraso parecido serão colocados, dessa forma consegue-se perceber clientes com mais ou menos atraso e isso pode influenciar na hora da classificação.
- Então estes clientes desses subgrupos são submetidos a um cálculo para realizar avaliação, neste cálculo vai haver históricos de pagamento, renda, profissão, cobranças anteriores, situação financeira entre outros.
- Após este cálculo será dada uma nota de 0 a 1000 que vai ajudar a otimizar a recuperação, quanto mais próximo de 1000 maiores as chances de o credor recuperar o crédito solicitado pelo cliente.

Tendo estes dados é possível realizar a personalização dos dados apresentados, pois clientes com maior pontuação e com maior chance de realizarem o pagamento, podem receber maiores descontos e assim firmar essa parceria entre empresa e cliente (SOUSA, 2018).

Além disso, o *Collection Score* não serve apenas para recuperação, mas pode auxiliar na hora de fazer a captação do cliente, pois tendo um histórico do cliente, fica muito mais fácil oferecer crédito para aqueles que sempre cumprem com seus contratos. Com essa técnica as cobranças ficam mais assertivas e isso faz com que o fluxo de caixa gire mais rápido tornando o processo mais rápido e eficaz.

### 1.3 Aprendizagem de Máquina

Quando Arthur Samuel (1959) criou o termo *Machine Learning* ele tinha o conceito que seria “um campo de estudo que dá aos computadores a habilidade de aprender sem terem sido programados para tal”, mas com os estudos na área e com as novas tecnologias que vem surgindo, o aprendizado de máquina vem evoluindo cada vez mais, fazendo parte de nos dias de formas que muitas vezes não percebemos (GOMES, 2019).

A aprendizagem de máquina é uma subárea da inteligência artificial que tem como principal objetivo treinar as máquinas para que aprendam com os dados disponíveis. Com a ascensão da internet, a quantidade de dados gerenciados cresceu exponencialmente e a possibilidade de uso desses dados nos negócios ficou cada vez mais evidente, pois esses tipos de algoritmos podem prever e guiar uma tomada de decisão a partir de amostras de dados (LUDERMIR, 2021).

Essa tomada de decisão faz com que haja diversas inovações nos mercados, de forma que a aprendizagem de máquina pode ser utilizada para reconhecimento de voz, reconhecimento de imagens, sugestões de filmes, séries e músicas, automação de carros e casas entre outros serviços que podem fazer uso dessa tecnologia (SILVA, 2020).

Mas para que estes algoritmos sejam os mais assertivos possíveis, é necessário que o algoritmo passe por um processo de aprendizagem que vai envolver o treinamento destes processos e dentro do aprendizado temos duas abordagens, o aprendizado supervisionado e o aprendizado não-supervisionado, dentro do aprendizado supervisionado encontra-se a classificação multirrótulo que é abordada nesse trabalho.

#### 1.3.1 Aprendizado Supervisionado

A partir de uma saída já conhecida é possível classificar novas informações a partir do que está sendo inserido na entrada, assim é como o aprendizado supervisionado funciona (BARROS, 2016). Neste tipo de aprendizado é fornecida uma referência de onde se deseja chegar, ou seja, recebe um conhecimento prévio do ambiente, contendo entradas e uma saída esperada.

Segundo Mitchell (1997) é possível calcular o erro gerado por uma aprendizagem supervisionada a partir da subtração da saída desejada e da saída gerada, para isso é utilizado o estímulo sofrido, o sinal de erro, a saída desejada e a saída gerada pelo algoritmo. O Objetivo



é que o algoritmo seja capaz de identificar e gerar saídas corretas a partir do que ele já aprendeu em treinamentos anteriores (PELLUCCI, et. al (2011)).

Este tipo de abordagem tenta prever os resultados da saída de acordo com a entrada, mas sabendo qual vai ser o tipo de saída gerada, além disso ela tenta resolver problemas como regressão e classificação. No caso da regressão é possível prever resultados para uma saída contínua, já para a classificação essa saída será para categorias diferentes de acordo com as características do dado analisado.

### 1.3.2 Classificação Multirrótulo

A classificação multirrótulo é o processo de associar diversos rótulos de uma classe para um dado usando suas características, existem diversas aplicações para este tipo de classificação nas quais incluem: classificação de artigos de jornais, genômica funcional, classificação semântica, classificação de músicas, predição do efeito colateral de drogas, categorização dos tipos de falhas de execução de programas a partir de análises de *crash reports*, atribuição automática dos gêneros de filmes e series entre outros tipos que podem ser estudados (GONCALVES, 2018).

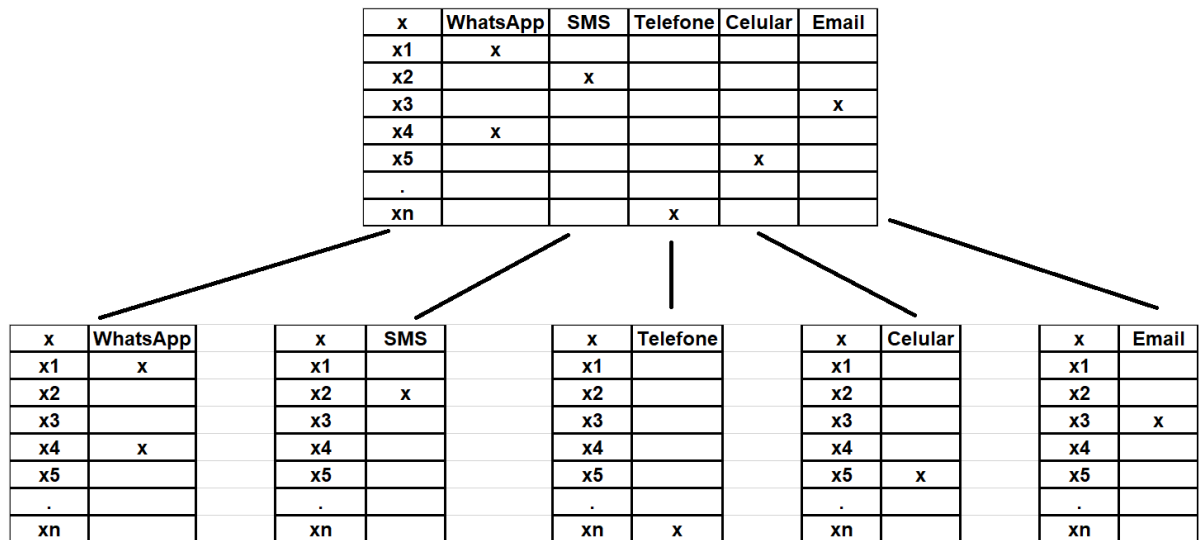
Em uma classificação multirrótulo, o exemplo pode ser associado a duas ou mais classes ao mesmo tempo, um classificador multirrótulo pode ser descrito com a seguinte função  $H: x \rightarrow 2^L$  que pega o exemplo  $x$  em um conjunto  $C \in 2^L$ , onde  $2^L$  é o conjunto potência de  $L$ , ou seja, o conjunto formado por todos os subconjuntos de  $L$ . No início a motivação para o estudo de multirrótulos era para tarefas de classificação de textos e diagnósticos médicos, mas com o passar do tempo e com os estudos muitas aplicações foram desenvolvidas, apesar da classificação de textos ser a mais estudada, pode-se encontrar trabalhos fazendo uso de multirrótulo com bioinformática, diagnóstico médico e classificação de imagens (CERRI, 2010).

No exemplo da classificação de textos pode-se exemplificar com por exemplo gêneros musicais, por exemplo: o gênero rock tem diversas ramificações dentro dele, como por exemplo o rock clássico, e temos também o gênero clássico, se tivermos uma base multirrótulo podemos colocar o rock, rock clássico e o clássico marcados como multirrótulo para uma determinada categoria de música e assim identificar em qual gênero aquela música pertence.

Alguns dos algoritmos de classificação multirrótulo que podem ser utilizados são o *Binary Relevance (BR)*, *Classifier Chain (CC)* e *Label Powerset (LP)*.

O BR funciona dividindo a base original em  $q$  bases de dados binárias para cada rótulo, assim fazendo a classificação em cada uma dessas novas bases binárias. Após o treinamento a classificação se dá de forma muito mais simples, pois quando um novo dado é apresentado, ele será predito com base nas combinações de saídas geradas em cada base binária (GONCALVES, 2018).

Figura 2 - Exemplo de decomposição Binária do classificador BR



Fonte: Autor (2022)

Este tipo de estratégia é vantajoso, pois é um método simples e é capaz de prever rótulos que não estão nas bases, além de apresentar uma complexidade linear em função de  $q$ , apesar disso, como os classificadores binários tomam decisões sem verificar outros rótulos, acabam ignorando relações de outros rótulos. O método BR apresenta bons resultados se comparados a outros métodos (SANTOS e ROSSI, 2020).

O pseudocódigo do BR funciona da seguinte forma:

*Inputs:*

1. Base de dados para Treino:  $TD = \{td1, td2, td3, td4, td5\}$
2. Conjunto Classificador:  $C = \{c_1, c_2, c_3, c_4, c_5 : c_i \in \{SVM, KNN, NB\}\}$
3. Rotulo:  $L \{Arts, Sport, Politics, Economy, and Science\}$
4. Conjunto de Teste

*Saida: MCR Resultado da classificação Multirrótulo com o conjunto de teste.*

---

For ( $i = 0; i < tamanho_{do\_dominio(s)}; i++$ )

*Início*

$Base_{Treino}(c_i, td_i)$  base de treinamento  $c_i$  usando domínio  $td_i$

*fim*

*Foreach*( document  $d$  in Conjunto de Teste)

*Início*

$document_{label} = \{\text{vazio}\};$

*Foreach*( Classificador  $C_i$  in  $C$ )

*Início*

$r_i = \text{Classify}(c_i, d)$  (resultado de  $c_i$  é 1 se o documento  $d$  pertencer a  $td_i$  ou 0 se não)  
se ( $r_1 == 1$ )

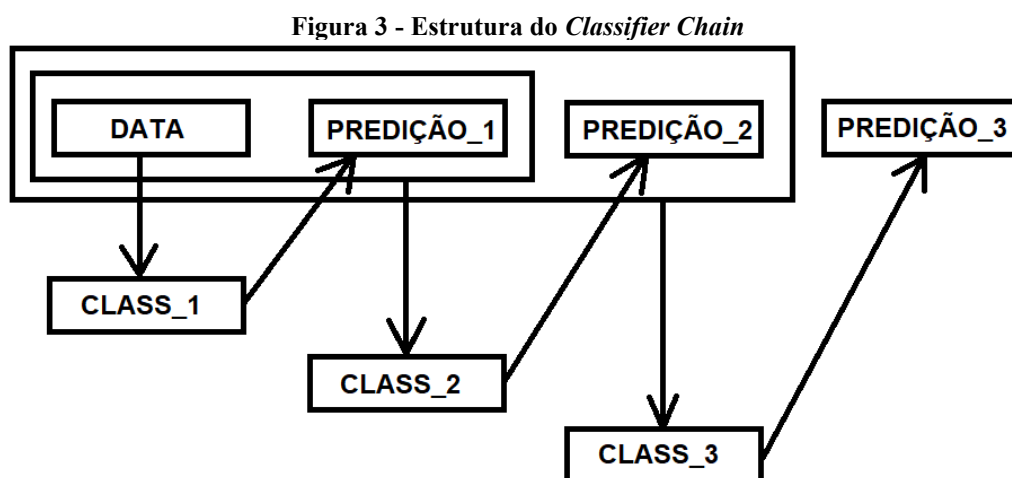
$document_{label}.add(label_i)$

*fim*

$MCL[d] = document_{label};$

*fim*

A abordagem CC é muito parecida com a BR, pois também utiliza classificadores binários em seu aprendizado. Neste tipo de abordagem, são gerados classificadores binários conectados em uma cadeia, sendo que novos classificadores são ligados a partir de um novo rótulo previsto pelo classificador anterior, então suas previsões levam em conta o que classificadores anteriores geraram para gerar novas previsões e assim por diante (SANTOS e ROSSI, 2020) assim como mostra a Figura 3.



Fonte: Autor (2022)

O pseudocódigo do CC é separado em duas funções, uma de treinamento e outra de classificação e funciona da seguinte maneira:

```

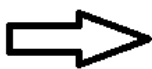
Trainamento( $D = \{(X_1, S_1), \dots, (X_n, S_n)\}$ )
  For  $j \in 1 \dots |L|$ 
    do single – label transformação e treinamento
     $D' \leftarrow \{ \}$ 
    For  $(x, S) \in D$ 
      do  $D' \leftarrow D' \cup (x, l_1, \dots, l_j - 1), l_j)$ 
    train  $C_j$  para prever a relevancia binaria de  $l_j$ 
     $C_j: D' \rightarrow l_j \in \{0,1\}$ 

  Classificar( $x$ )
     $Y \leftarrow \{ \}$ 
    For  $j \leftarrow 1$  até  $|L|$ 
      do  $Y \leftarrow Y \cup (l_j \leftarrow C_j : (x, l_1, \dots, l_j - 1))$ 
    return  $(x, Y)$  exemplo de classificação
  
```

O LP é uma abordagem de transformação de um problema multirrótulo em um problema multiclasse. Para cada conjunto de rótulos é considerado um conjunto de treinamento multirrótulo que vai transformar rótulos multirrótulo em rótulos simples, ou seja, cada rótulo é considerado único. Um texto rotulado como “Economia” e “Esporte” vai ter um novo rótulo chamado “Economia-Esporte” após a aplicação do método (GONCALVES, 2018).

Figura 4 - Estrutura do Label Powerset

Ex#	Label set
1	(a1, a4)
2	(a3, a4)
3	(a1)
4	(a2, a3, a4)



Ex#	Label
1	1001
2	1100
3	1000
4	0111

Fonte: Autor (2022)

Esse tipo de abordagem pode aumentar o tempo de aprendizagem dependendo do algoritmo, principalmente em algoritmos de classificação binário. Um dos maiores problemas com o LP é que algumas combinações podem não ser tão frequentes, fazendo que isso afete negativamente a classificação (HENRIQUE e ROSSI, 2020).

#### 1.4 Sistemas de Recomendação

Com a evolução e disseminação da tecnologia as informações geradas são cada vez mais precisas, isso porque a todo momento estamos gerando dados e estes dados são analisados e devolvidos para o usuário de alguma forma, seja em forma de propagandas, produtos, músicas, filmes, viagens, amigos, lugares para ir entre outras formas de mostrar conteúdo personalizado para cada usuário, e isso se dá pelo fato da existência dos sistemas de recomendação.

Os sistemas de recomendação são ferramentas que realizam a interpretação dos dados, para então auxiliar no processo de exibir para o usuário um conteúdo personalizado de acordo com as preferências do mesmo, estes perfis são montados de acordo com avaliações que o usuário realizou com base no seu interesse, seja curtindo algo, compartilhando, visualizando diversas vezes um determinado produto ou então realizando uma pesquisa (CARVALHO e MACEDO, 2014).

Esta área começou por volta dos anos 90 e um dos primeiros sistemas criado foi o RINGO, que com informações de perfis de usuários recomendava músicas, na mesma época outro sistema foi criado para recomendação de notícias o GroupLens, mais tarde o sistema MovieLens de recomendação de filmes foi criado, estes sistemas foram a base do que conhecemos de sistemas de recomendação atualmente (MOTTA, 2011). Estes sistemas são baseados no que é comum para os usuários, então se algo é interessante para um usuário com um perfil, então um conteúdo parecido pode ser apresentado para um outro usuário com o perfil parecido.

Muitas empresas são conhecidas por utilizarem dessa técnica para poderem obter maior lucro e fazerem com que o conteúdo seja personalizado de acordo com cada indivíduo, como exemplo pode-se citar a Amazon. A Amazon é referência em quando se fala em recomendação, seu algoritmo *Item-to-Item Collaborative Filtering* funciona da seguinte forma: ao selecionar um produto em seu site, outros produtos são mostrados como sugestão, pois outros usuários que compraram o primeiro também compraram um segundo que está sendo sugerido (TAKAHASHI, 2015). Um outro exemplo é a Netflix, que ao selecionar filmes de um certo ator ou então gênero, outros filmes com a mesma temática ou ator são

recomendados. Isso vale para o Spotify e Youtube que ao selecionar uma música consegue recomendar a próxima música ou vídeo que será reproduzido baseado no histórico de reprodução e nos estilos escolhidos.

Esse tipo de personalização enfrenta um problema chamado Cold Start, esse termo é de uso da área automobilística, mas ganhou uso na recomendação pois trata-se do início de uma recomendação sem conhecer o usuário para qual o sistema deve recomendar (PINHEIRO, 2019). O sistema irá verificar o usuário e perceber que não existem informações suficientes para que a recomendação aconteça de maneira que o conteúdo seja personalizado para aquele perfil, dessa forma o sistema tende a mostrar conteúdos que as pessoas mais acessam, ou seja, conteúdos em alta. Entre as abordagens que se propõem a resolver esse problema está a aprendizagem ativa, pois a partir do momento que o usuário começa a interagir com o sistema, os dados começam a ser gerados e é possível personalizar as recomendações para ele.

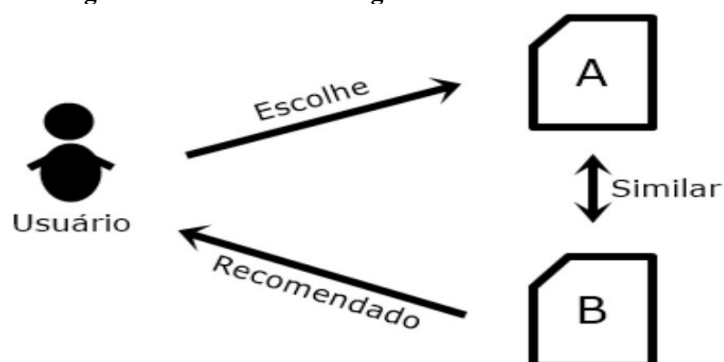
Tudo isso é possível graças aos métodos de predição utilizados por esses algoritmos, e os sistemas de recomendação podem ser classificados em três categorias propostas por Shoham e Balabanovic (1997), são elas baseadas em conteúdo, colaborativas e híbridas.

#### 1.4.1 Filtragem Baseada em Conteúdo

A abordagem de filtragem baseada em conteúdo se baseia em identificar os padrões de conteúdo ou itens desejados e conseguir recomendar itens semelhantes que o usuário poderá gostar também, esta identificação se dá por meio de cálculos matemáticos baseados nas propriedades dos dados, e isso faz com que o sistema identifique que um item é semelhante ao outro (BONIN, 2018).

Este sistema atua com base no reforço, ou seja, quanto mais informações os usuários passam em relação a um determinado produto, mais preciso ele fica. Por exemplo quando é pesquisado um termo em um site de busca, conforme são selecionados resultados de acordo com o que é pesquisado, mais preciso o algoritmo fica para em uma próxima busca trazer aqueles resultados que foram mais selecionados e deixa de trazer o que for menos relevante (D'ADDIO, 2015).

Figura 5 - Modelo de Filtragem Baseada em Conteúdo



Fonte: Rolim et.al (2018)

Os sistemas de recomendação baseados em conteúdo contêm as seguintes vantagens (SILVA, 2014):

- Não é necessário conhecer o usuário.
- Recomendações personalizadas de acordo com o perfil de cada usuário, independentemente dos gostos.
- O algoritmo pode se adaptar a novos itens.
- Pode disponibilizar informações sobre a recomendação.

Apesar de ser um sistema muito utilizado também existem desvantagens, são elas (SILVA, 2014)

- Necessidade de um grande treinamento para encontrar características apropriadas.
- Só são recomendados itens dentro do perfil do usuário.
- As filtragens são feitas de acordo com as notas dos usuários, ou seja, nem sempre podem ser precisas pela qualidade ou serventia dos itens.

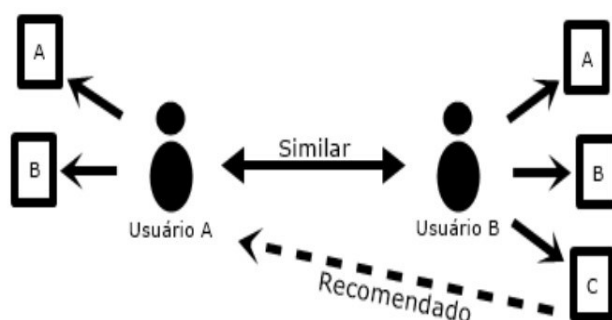
#### 1.4.2 Filtragem Colaborativa

Esta técnica é baseada nos gostos de usuário distintos, isso faz com que diferente a técnica de filtragem baseada em conteúdo, não seja necessário conhecer o item analisado, mas sim as preferências dos usuários, e assim recomendar de acordo com os gostos similares (D'ADDIO, 2015).

Por exemplo, caso um usuário 1 goste das músicas A, B e C, e o usuário 2 goste das músicas B, C e D, então o usuário 1 receberá uma recomendação da música D e o usuário 2 da música A (BALDÁRRAGO, 2012).

É um dos algoritmos mais utilizados da recomendação, pois consegue aprender melhor e para isso existem dois tipos de algoritmo que podem ser utilizados: a filtragem colaborativa usuário - usuário e a item - item.

**Figura 6 - Exemplo de Recomendação baseada filtragem colaborativa**



**Fonte: Rolim et.al (2018)**

- Filtragem colaborativa usuário – usuário: São encontradas similaridades entre os usuários de acordo com os gostos que eles escolheram em um passado. O algoritmo demanda tempo para aprender, mas é muito eficiente.
- Filtragem colaborativa item – item: Muito parecido com o anterior, mas ao invés de tentar encontrar semelhanças entre usuários, as semelhanças são encontradas nos itens escolhidos pelos usuários.

Nos dois casos é possível determinar quão similares são os usuários ou os itens, mas este tipo de abordagem também tem seus problemas e eles podem aparecer quando:

- São inseridos novos usuários ou itens.
- Os itens avaliados não são encontrados em outros usuários.
- Tem tendência a recomendar itens populares, sejam eles bons ou ruins.

### 1.4.3 Filtragem Híbrida

Como apresentado anteriormente existem alguns problemas nas abordagens baseadas em conteúdo e colaborativas, a filtragem híbrida foi criada para utilizar das vantagens de cada



uma dessas abordagens e tentar diminuir a quantidade de problemas que podem surgir. Este tipo de filtragem faz uso das duas abordagens e normalmente o sistema vai realizar uma filtragem colaborativa e logo após uma filtragem baseada em conteúdo (BARBOSA, 2014).

Serviços de streaming utilizam esse tipo de abordagem, pois ao entrarmos no seu catálogo é nos indicado um filme baseado no que as pessoas recomendaram, e logo após assistir um filme da lista o próximo filme recomendado vai ser de acordo com filmes assistidos anteriormente pelo usuário (LIMA, 2012).

Para realizar esse tipo de abordagem são ajustados pesos para que a recomendação fique mais precisa, são eles *Weighter*, *Switching*, *Mixed*, *Feature Combination*, *Cascade* e *Meta-Level*.

- *Weighted*: Este modelo de peso dá uma nota baseada em todas as técnicas que o sistema utiliza, a vantagem é que o peso é ajustado de acordo com qual técnica é utilizada.
- *Switching*: Trabalha com alternância de modelos utilizados, busca evitar os problemas de um modelo trocando a técnica utilizada para um outro e complementando a recomendação.
- *Mixed*: Este modelo é recomendado quando tem que se fazer diversas recomendações em paralelo, este modelo reduz os problemas que surgem ao surgir um novo item.
- *Feature Combination*: Aqui uma técnica é utilizada para coletar características que serão utilizadas na próxima técnica, assim os dados processados são distintos.
- *Cascade*: O modelo cascade envolve estágios de recomendação, pois haverá uma recomendação intermediária e a partir dela, uma segunda técnica é aplicada para refinar a recomendação.
- *Meta-Level*: Por fim são combinadas duas técnicas, onde uma servirá de entrada para outra técnica utilizada.

Das três técnicas apresentadas, a filtragem híbrida é a que contém uma análise de dados mais completa, e acaba sendo uma ótima escolha a aplicação desta abordagem na construção de algoritmos de recomendação, pois possibilita realizar recomendações precisas e variadas (AGNOLO, MOURA e CUNHA, 2021), as precisas serão baseadas em conteúdo e as variadas em usuários. As vantagens apresentadas são:

- Recomendações excelentes para itens parecidos.

- Tem a possibilidade de variar os tipos de recomendações, pois como utiliza duas ou mais técnicas, o usuário sempre terá recomendações distintas.

## 1.5 Métricas para Avaliação

A construção de um modelo de recomendação envolve diversos fatores, desde escolher qual filtragem será utilizada até mesmo conseguir avaliar o quão preciso o sistema é, para isso temos as métricas de avaliação (BARBOSA, 2014).

A precisão, Figura 7, é medida através da ordem do que é mais relevante o que não. É mostrada como a razão dos itens recomendados interessantes para o usuário pela quantidade total de itens recomendados.

**Figura 7 - Equação de Precisão**

$$\text{Precisão} = \frac{VP}{VP + FP} = \frac{VP}{REC}$$

**Fonte: Barbosa (2014)**

O Recall, Figura 8, pode recuperar os itens que se mostram relevantes para o sistema, e indica o total de itens recomendados para o usuário, que pode ser chamado de taxa de verdadeiros positivos. É utilizada para mostrar a relação entre previsões positivas encontradas e todas as outras previsões positivas.

**Figura 8 - Definição do cálculo do recall**

$$\text{Recall} = \frac{VP}{VP + FN} = \frac{VP}{REL}$$

**Fonte: Barbosa (2014)**

A *F-measure* ou  $F_1$ , Figura 9, vai combinar tanto a precisão quanto o recall, é chamado de média harmônica e está relacionada a grandezas inversamente proporcionais, o resultado é parecido quando com a média aritmética quando os valores são próximos (BARBOSA, 2014).

**Figura 9 - Cálculo do F-measure**

$$F1 = \frac{2 \cdot \text{Precisão} \cdot \text{Recall}}{\text{Precisão} + \text{Recall}}$$

**Fonte: Barbosa (2014)**

## 1.6 Trabalhos Similares

Para o desenvolvimento deste trabalho foram analisadas monografias, dissertação de mestrado, artigos científicos, sites entre outras fontes para contextualização utilizando para busca as palavras-chave: Sistemas de Recomendação, Dividas, Finanças.

Entre estes trabalhos pode-se destacar o trabalho realizado por Forti (2018) que em sua dissertação demonstrou técnicas de *machine learning* no mercado de crédito brasileiro, o que se assemelha com o trabalho aqui desenvolvido, pois utiliza o método do *collection score* para definir os clientes mais propensos a pagar, mas no caso de seu trabalho ela buscou realizar uma comparação entre metodologias tradicionais de regressão logística e o *machine learning* para o mercado de crédito.

Em seu trabalho Forti demonstrou que utilizando o método Gradient Boosting ela poderia obter uma melhor classificação em comparação com outros métodos estudados, apesar de durante suas conclusões demonstrar também que é necessário um ambiente preparado por conta do alto número de regras.

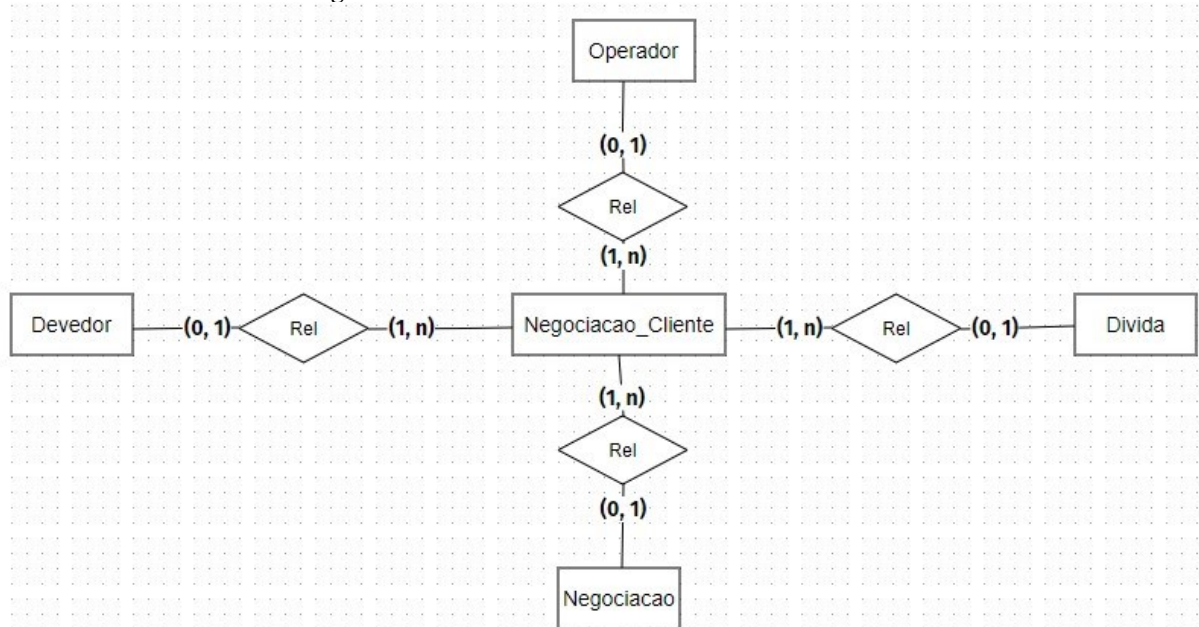
Um outro trabalho que se pode destacar é o de Lauria (2020) que em sua dissertação apresentou uma revisão bibliográfica sobre as aplicações de sistemas de recomendação em finanças, realizando recomendações em plataformas digitais de investimento.

### 3. DESENVOLVIMENTO

Para início de desenvolvimento foi necessária a criação de uma base de dados sintética, onde todos os seus dados e estrutura foram baseados em uma base real de dívidas, mas para preservar as informações sensíveis e não ir contra a LGPD (Lei Geral de Proteção de Dados) optou-se pela versão sintética. Para isso, analisando a base real, foi desenvolvido um modelo relacional pensado em como os dados se conectam, apresentado na Figura 10.

Também, adotou-se o modelo de recomendação colaborativo baseado em usuário, visto que as informações de dívidas pagas de outros usuários é que irão contribuir para recomendar ações através do classificador multirrótulo para um usuário devedor.

**Figura 10 - Modelo Relacional da base de Devedores**



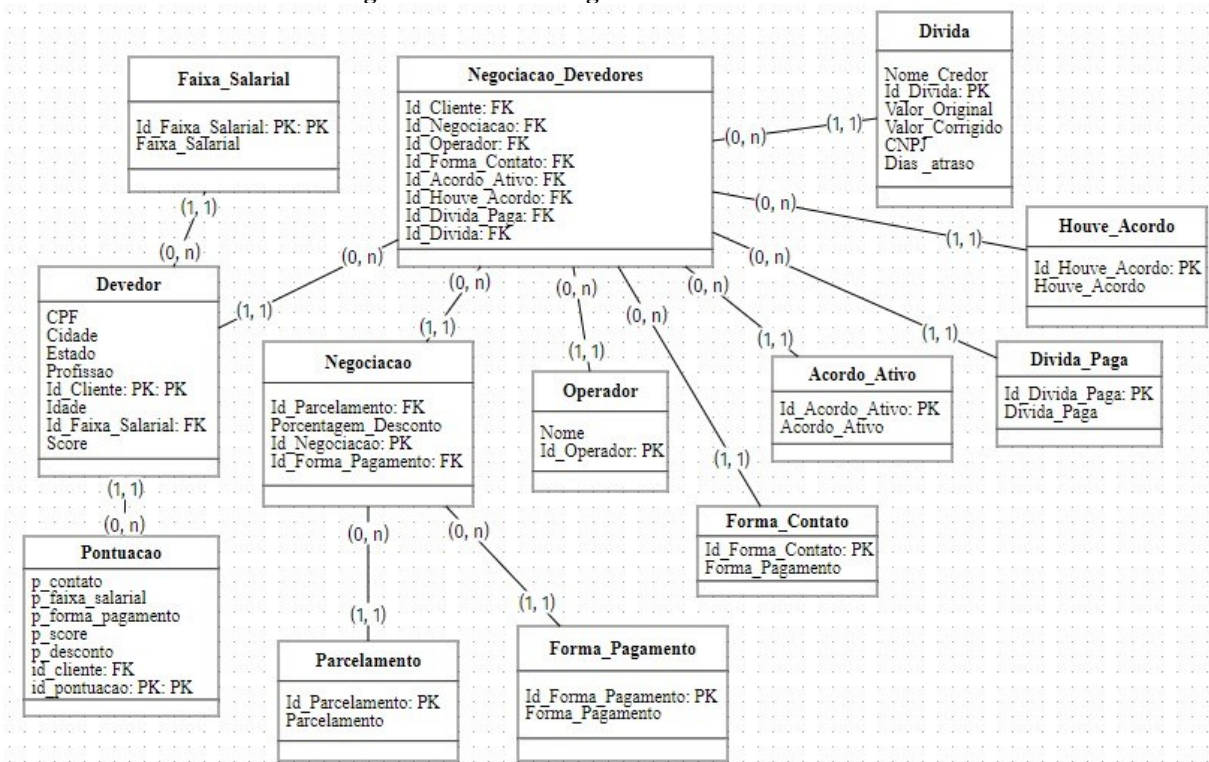
Fonte: Autor (2022)

Tendo gerado o modelo relacional, foi criado um modelo lógico, onde analisando como alguns dados apareceriam na base, foram adicionadas novas tabelas no projeto para que estas comportassem dados de identificação da coluna, fazendo assim que em algumas tabelas existissem apenas dados de identificação.

Estas novas tabelas foram “Faixa\_Salarial”, “Parcelamento”, “Forma\_Pagamento”, “Forma\_Contato”, “Divida\_Paga”, “Houve\_Acordo” e “Acordo\_Ativo”.

A tabela “Negociacao\_Devedores” foi pensada para ser a tabela que vai conectar todas as outras, sendo ela a tabela responsável por gerar os dados finais que serão utilizados no modelo gerado está na Figura 11.

Figura 11 - Modelo lógico da base de devedores



Fonte: Autor (2022)

Tendo o modelo lógico definido, foi gerado um modelo físico utilizando o SQL Server e então populado com os dados gerados de forma aleatória de acordo com a base real.

Estes dados foram gerados aleatoriamente ajustando alguns dados por meio de um algoritmo desenvolvido em C#. Neste algoritmo foram criadas funções que retornam *strings* de inserção na base. Foram criadas cinco funções com os seguintes retornos, respectivamente: Devedores, Negociação, Operadores, Dívida e Negociação Devedores.

Devedores recebem os seguintes dados: CPF, Cidade, Estado, Profissão, Faixa\_Salarial, Idade e Score.

A *string* de Negociação recebe os seguintes dados: Id\_Forma\_Pagamento, Id\_Parcimento e porcentagem\_desconto, todos estes dados são gerados aleatoriamente.

Operadores gera uma *string* de inserção com dez nomes aleatórios para utilizarmos como se fosse o operador de cobrança.

Dívida vai gerar a *string* com Nome\_credor, valor\_original, valor\_corrigido, CNPJ e dias\_atraso. Na criação da *string* de Negociacao\_Devedores serão gerados os seguintes dados: Id\_Cliente, id\_Divida, Id\_operador, id\_forma\_contato, id\_acordo\_ativo, Id\_houve\_acordo e id\_divida\_paga.

No total foram gerados 1000 registros de dívidas para inserção na base sintética, estes dados após gerados passaram por uma filtragem para identificar quais seriam os melhores dados que poderiam fazer parte da classificação multirrótulo, então foram removidas algumas colunas para minimizar o impacto que dados com tipos de texto podem ocasionar na hora de realizar a classificação. O *script* gerado para selecionar os dados para a tabela de devedores está descrito na Figura 12.

**Figura 12 - Script para seleção dos dados**

```
SELECT d.estado, d.idade, D.SCORE, di.nome_credor, di.Valor_Original
], ne.porcentagem_desconto, di.Valor_Corrigido, di.Dias_atraso, fa.Faixa_Salarial,
pa.parcelamento, fc.Forma_Contato, fp.Forma_Pagamento, dp.divida_paga from negociacao_devedores as n
inner join Devedores as d on d.Id_Cliente = n.id_cliente
inner join Divida as di on di.id_divida = n.id_divida
inner join Negociacao as ne on ne.id_negociacao = n.id_negociacao
inner join Operador as o on o.Id_Operador = n.Id_Operador
inner join Faixa_Salarial as fa on fa.Id_Faixa_Salarial = d.Id_Faixa_Salarial
inner join Houve_Acordo as ha on ha.Id_Houve_Acordo = n.Id_Houve_Acordo
inner join Parcelamento as pa on pa.Id_Parcelamento = ne.Id_Parcelamento
inner join Forma_Pagamento as fp on fp.Id_Forma_Pagamento = ne.Id_Forma_Pagamento
inner join Forma_Contato as fc on fc.Id_Forma_Contato = n.Id_Forma_Contato
inner join Acordo_Ativo as aa on aa.Id_Acordo_Ativo = n.Id_Acordo_Ativo
inner join Divida_Paga as dp on dp.Id_Divida_Paga = n.Id_Divida_Paga
```

**Autoria: Própria**

Para o início do desenvolvimento foram criadas algumas regras para que na geração dos modelos seguissem alguns padrões estabelecidos.

A primeira regra é a quantidade de dias em atraso, pois como este trabalho vai tratar apenas dos casos de *Collection Score*, então foi restrita uma faixa de atraso entre 15 e 180 dias, também foram definidas formas de contato baseadas nas reais, são elas “WhatsApp”, “Celular”, “Telefone”, “SMS” e “Email” conforme o Quadro 1.

**Quadro 1 - Dados da Tabela Forma de Contato**

Id_Forma_Contato	Forma_Contato
1	Celular
2	WhatsApp
3	Telefone
4	SMS
5	Email

**Fonte: Autor (2022)**

As formas de pagamento foram definidas como sendo “Cartão”, “Boleto”, “Pix”, “Picpay” e “DÉBITO AUTOMÁTICO” conforme Quadro 2.

**Quadro 2 - Dados da Tabela Forma de Pagamento**

Id_Forma_Pagamento	Forma_Pagamento
1	DÉBITO AUTOMÁTICO
2	CARTAO
3	PIX
4	PICPAY
5	BOLETO

**Fonte: Autor (2022)**

Os valores das tabelas “Houve\_Acordo”, “Parcelamento”, “Acordo\_Ativo” e “Divida\_Paga” foram definidos sendo respectivamente “SIM” e “NÃO”. Após as definições de valores, foi gerada a base e uma visualização da base conforme Quadro 3.

Quadro 3 - Visualização contendo 10 dados gerados para a base de devedores no SQL Server

UF	Idade	Score	Nome_Credor	Porcentagem_De sconto	Valor_Orig inal	Dias_Atr aso	Valor_Corri gido	Faixa_Sala rial	Parcelame nto	Forma_Con tato	Forma_Paga mento	Divida_P aga
AM	38	442	BV FINANCEIR A	31	542	55	840,1	1001 a 2000	SIM	Email	PIX	SIM
RJ	29	249	BV FINANCEIR A	55	5086	28	6510,08	Menos de 1000	NAO	WhatsApp	BOLETO	NAO
PR	25	496	CLARO NET	36	9044	37	12390,28	Menos de 1000	SIM	Celular	PICPAY	NAO
MG	50	725	BRADESCO	11	4977	109	10401,93	Menos de 1000	SIM	Email	DÉBITO AUTOMAT ICO	NAO
GO	54	429	BRADESCO	11	9291	18	10963,38	6001 a 7000	SIM	Celular	PIX	NAO
BA	49	836	MAPFRE SEGUROS	82	1152	101	2315,52	1001 a 2000	SIM	SMS	PIX	NAO
PI	52	158	SAFRA	50	5690	148	14111,2	4001 a 5000	SIM	SMS	BOLETO	NAO
BA	26	66	YAMAHA	3	6945	124	15556,8	2001 a 3000	NAO	WhatsApp	BOLETO	NAO



M G	22	5	VOLKSWA GEN	38	2915	81	5276,15	3001 a 4000	SIM	SMS	PIX	SIM
SP	35	704	SAFRA	81	3581	97	7054,57	1001 a 2000	SIM	Celular	CARTAO	NAO

Fonte: Autor (2022)

Como objetivo do trabalho é recomendar várias ações que podem ocorrer simultaneamente, a base foi transformada em multirrótulo, sendo cada rótulo uma ação que a equipe de cobrança pode tomar, então os dados gerados da visualização foram importados em um arquivo CSV onde passaram por uma nova filtragem e ajustados alguns valores para que se possa trabalhar com multirrótulo.

Após, foram criadas dois conjuntos de dados multirrrotulados, um onde cada linha das tabelas de “Porcentagem\_Desconto” e “Forma\_Pagamento” foram transformados em multirrótulos, Conjunto A, e outro onde as linhas das tabelas “Porcentagem\_Desconto”, “Forma\_Pagamento” e “Forma\_Contato” foram transformados, Conjunto B, assim será possível uma comparação de desempenho com mais ou menos rótulos.

A tabela “Porcentagem\_Desconto” foi separada em dez rótulos com faixas e cada uma dessas faixas pode assumir o valor 0 ou 1, os rótulos gerados foram “0\_A\_10”, “11\_A\_20”, “21\_A\_30”, “31\_A\_40”, “41\_A\_50”, “51\_A\_60”, “61\_A\_70”, “71\_A\_80”, “81\_A\_90” e “91\_A\_99”.

Isso foi feito para “Forma\_Pagamento”, que agora está representada pelos seguintes rótulos: “PIX”, “CARTAO”, “BOLETO” e “PICPAY”.

O mesmo procedimento foi feito para “Forma\_Contato” e os seguintes rótulos foram gerados: “WHATSAPP”, “SMS”, “CELULAR”, “TELEFONE” e “EMAIL”.

Feitos os ajustes necessários para utilizar a abordagem multirrótulo, esses dados foram colocados em um arquivo com extensão ARFF, Figura 13, e ele foi submetido a ferramenta MEKA, pois esta oferece um ambiente com diversos algoritmos de classificação multirrótulo implementados e com manuseio através de interface gráfica.

**Figura 13 - Exemplo de arquivo com extensão ARFF**

```
@relation DividasDevedores
@attribute UF {'AC','AL','AP','AM','BA','CE','ES','GO','MA','MT','MS',
'MG','PA','PB','PR','PE','PI','RJ','RN','RS','RO','RR','SC','SP','SE','TO','DF'}
@attribute IDADE numeric
@attribute BANCO {'HONDA','ITAU','BRADESCO','VOLKSWAGEN','PAN','BANCO DO BRASIL',
'SANTANDER','SAFRA','BV FINANCEIRA','MAPFRE SEGUROS','CLARO NET','RENNER','AVON','YAMAHA'}
@attribute VALORINICIAL numeric
@attribute DESCONTO numeric
@attribute 0_A_10 {0,1}
@attribute 11_A_20 {0,1}
@attribute 21_A_30 {0,1}
@attribute 31_A_40 {0,1}
@attribute 41_A_50 {0,1}
@attribute 51_A_60 {0,1}
@attribute 61_A_70 {0,1}
@attribute 71_A_80 {0,1}
@attribute 81_A_90 {0,1}
@attribute 91_A_99 {0,1}
@attribute VALORCORRIGIDO numeric
@attribute DIASATRASSO numeric
@attribute FAIXASALARIAL {'Menos de 1000', '1001 a 2000', '2001 a 3000', '3001 a 4000',
'4001 a 5000', '5001 a 6000', '6001 a 7000', 'acima de 7001'}
@attribute ACORDOATIVO {'SIM','NAO'}
@attribute WHATSAPP {0,1}
@attribute SMS {0,1}
@attribute CELULAR {0,1}
@attribute TELEFONE {0,1}
@attribute EMAIL {0,1}
@attribute CARTAO {0,1}
@attribute PIX {0,1}
@attribute BOLETO {0,1}
@attribute PICPAY {0,1}
@attribute SCORE numeric
@attribute DIVIDAPAGA {'SIM','NAO'}
@data
'BA',34,'CLARO NET',2197,63,0,0,0,0,0,0,1,0,0,0,5887.96,168,'6001 a 7000','SIM',1,0,0,0,0,1,0,0,0,791,'SIM'
'MT',50,'AVON',10729,30,0,0,1,0,0,0,0,0,0,0,21350.71,99,'1001 a 2000','SIM',0,1,0,0,0,0,1,0,0,567,'SIM'
'RS',39,'RENNER',1090,92,0,0,0,0,0,0,0,0,1,2670.5,145,'2001 a 3000','SIM',0,1,0,0,0,1,0,0,0,569,'SIM'
```

Fonte: Autor (2022)

Para os experimentos foram testados três algoritmos de classificação multirrótulo que são o BR, CC e LC combinados com os algoritmos J48, *Naive Bayes* e *Multilayer Perceptron*, totalizando nove algoritmos de classificação multirrótulo a serem testados com cada uma das duas bases de dados sintéticas multirrótulo.

#### 4. RESULTADOS E DISCUSSÕES

Os testes foram feitos utilizando o MEKA e os seguintes parâmetros foram configurados:

- *Split Percentage* = 70.0
- *Threshold* (Limite para seleção de rótulo de saída) = 0.5
- Todos os algoritmos foram executados com as configurações padrão

Os algoritmos de classificação multirrotulo selecionados para os testes foram o BR, CC e o LC, combinados com o J48, *Naive Bayes* e o *Multilayer Perceptron* para cada algoritmo.

Os resultados apresentados utilizando o Conjunto A para as métricas do algoritmo BR estão dispostos na Tabela 1 e os resultados utilizando o Conjunto B estão na Tabela 2.

**Tabela 1 - Métricas de avaliação para o algoritmo BR/Conjunto A**

BR	J48	<i>NaiveBayes</i>	<i>MultilayerPerceptron</i>
<i>Accuracy</i>	0,668	0,384	0,262
<i>Micro Precision</i>	1	0,789	0,393
<i>Micro Recall</i>	0,601	0,383	0,355
<i>Macro Precision</i>	0,714	0,791	0,344
<i>Macro Recall</i>	0,714	0,406	0,321
<i>F1 (micro averaged)</i>	0,751	0,516	0,373
<i>F1 (macro averaged by example)</i>	0,779	0,44	0,323
<i>F1 (macro averaged by label)</i>	0,714	0,506	0,327

Fonte: Autor (2022)

**Tabela 2 - Métricas de avaliação para o algoritmo BR/Conjunto B**

BR	J48	<i>NaiveBayes</i>	<i>MultilayerPerceptron</i>
<i>Accuracy</i>	0,558	0,319	0,241
<i>Micro Precision</i>	1	0,759	0,354
<i>Micro Recall</i>	0,430	0,275	0,318
<i>Macro Precision</i>	0,526	0,646	0,311
<i>Macro Recall</i>	0,526	0,303	0,290

<i>F1 (micro averaged)</i>	0,601	0,404	0,335
<i>F1 (macro averaged by example)</i>	0,668	0,383	0,312
<i>F1 (macro averaged by label)</i>	0,526	0,384	0,294

Fonte: Autor (2022)

O segundo algoritmo testado foi o CC, e assim como no primeiro teste, todos os algoritmos utilizados no primeiro foram novamente utilizados no segundo, os resultados apresentados estão na Tabela 3 e Tabela 4.

**Tabela 3 - Métricas de avaliação para o algoritmo CC/Conjunto A**

CC	J48	<i>NaiveBayes</i>	<i>MultilayerPerceptron</i>
<i>Accuracy</i>	0,713	0,482	0,410
<i>Micro Precision</i>	0,741	0,691	0,450
<i>Micro Recall</i>	0,741	0,501	0,451
<i>Macro Precision</i>	0,764	0,755	0,449
<i>Macro Recall</i>	0,799	0,512	0,448
<i>F1 (micro averaged)</i>	0,741	0,581	0,450
<i>F1 (macro averaged by example)</i>	0,785	0,540	0,465
<i>F1 (macro averaged by label)</i>	0,777	0,578	0,438

Fonte: Autor (2022)

**Tabela 4 - Métricas de avaliação para o algoritmo CC/Conjunto B**

CC	J48	<i>NaiveBayes</i>	<i>MultilayerPerceptron</i>
<i>Accuracy</i>	0,593	0,405	0,359
<i>Micro Precision</i>	0,592	0,620	0,396
<i>Micro Recall</i>	0,592	0,397	0,397
<i>Macro Precision</i>	0,585	0,612	0,408
<i>Macro Recall</i>	0,636	0,414	0,392
<i>F1 (micro averaged)</i>	0,592	0,484	0,396
<i>F1 (macro averaged by example)</i>	0,683	0,482	0,427
<i>F1 (macro averaged by label)</i>	0,601	0,472	0,388

O último algoritmo testado foi o LC, os resultados observados estão na Tabela 5 e Tabela 6, contendo resultados para utilização do Conjunto A, além de resultados com o Conjunto B, respectivamente.

**Tabela 5 - Métricas de avaliação para o algoritmo LC/Conjunto A**

LC	J48	<i>NaiveBayes</i>	<i>MultilayerPerceptron</i>
<i>Accuracy</i>	0,710	0,668	0,276
<i>Micro Precision</i>	0,737	0,705	0,330
<i>Micro Recall</i>	0,737	0,707	0,329
<i>Macro Precision</i>	0,786	0,748	0,307
<i>Macro Recall</i>	0,787	0,747	0,302
<i>F1 (micro averaged)</i>	0,737	0,706	0,329
<i>F1 (macro averaged by example)</i>	0,780	0,739	0,322
<i>F1 (macro averaged by label)</i>	0,782	0,742	0,296

Fonte: Autor (2022)

**Tabela 6 - Métricas de avaliação para o algoritmo LC/Conjunto B**

LC	J48	<i>NaiveBayes</i>	<i>MultilayerPerceptron</i>
<i>Accuracy</i>	0,595	0,522	0,223
<i>Micro Precision</i>	0,592	0,542	0,266
<i>Micro Recall</i>	0,737	0,569	0,265
<i>Macro Precision</i>	0,786	0,584	0,237
<i>Macro Recall</i>	0,787	0,591	0,232
<i>F1 (micro averaged)</i>	0,737	0,555	0,265
<i>F1 (macro averaged by example)</i>	0,780	0,617	0,278
<i>F1 (macro averaged by label)</i>	0,782	0,585	0,230

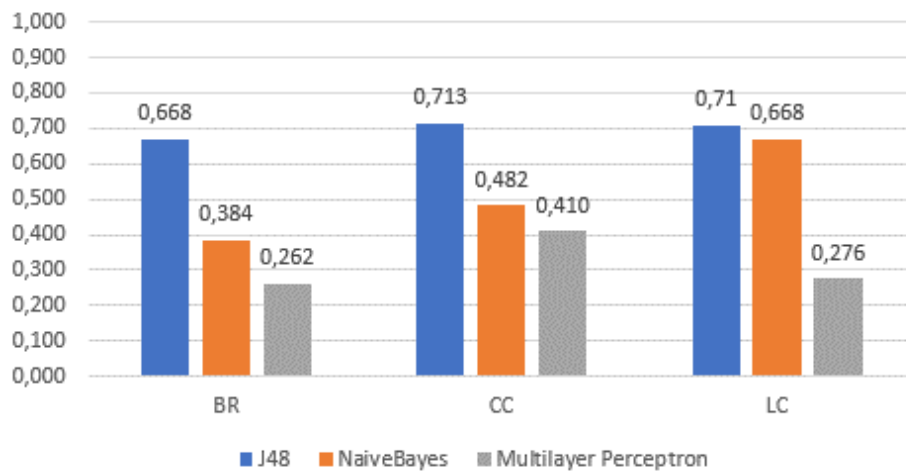
Ao realizar uma análise dos resultados obtidos, pode-se observar uma diferenciação dos resultados quando se faz uso de mais ou menos rótulos utilizando a classificação multirrótulo. Ao utilizar o Conjunto A, com menos rótulos, e os algoritmos propostos, foi possível detectar resultados superiores, avaliando as métricas, ao comparar com resultados obtidos ao utilizar o Conjunto B, com mais rótulos.

Isto demonstra que, ao aumentar a quantidade de rótulos os resultados vão se deteriorando, e, que o uso da classificação multirrótulo é pertinente para a recuperação de

crédito, já que a quantidade de ações recomendadas não é grande, mas em um cenário onde a quantidade de recomendações seja maior, os resultados podem não ser bons, ao considerar as métricas avaliadas.

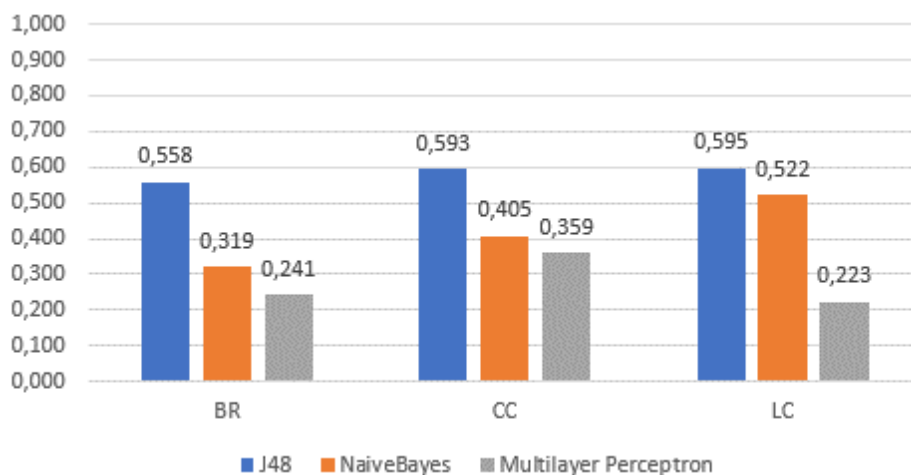
A partir destes resultados, também foi possível gerar gráficos mostrados como cada algoritmo se comportou em relação as métricas de avaliação de recomendação, estes resultados foram inseridos e mostrados nas Figuras 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 e 29:

**Figura 14 - Medida de Acurácia para o Conjunto A**



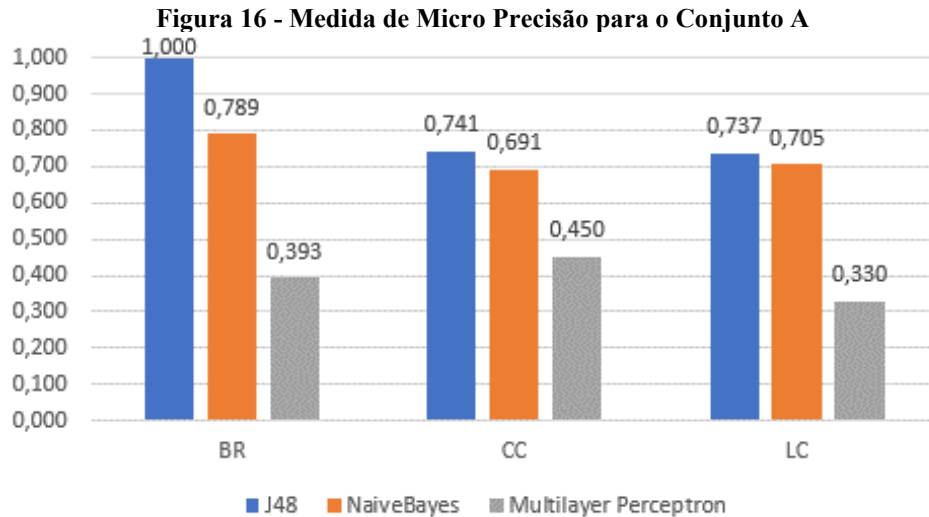
Fonte: Autor (2022)

**Figura 15 - Métrica de Acurácia para o Conjunto B**

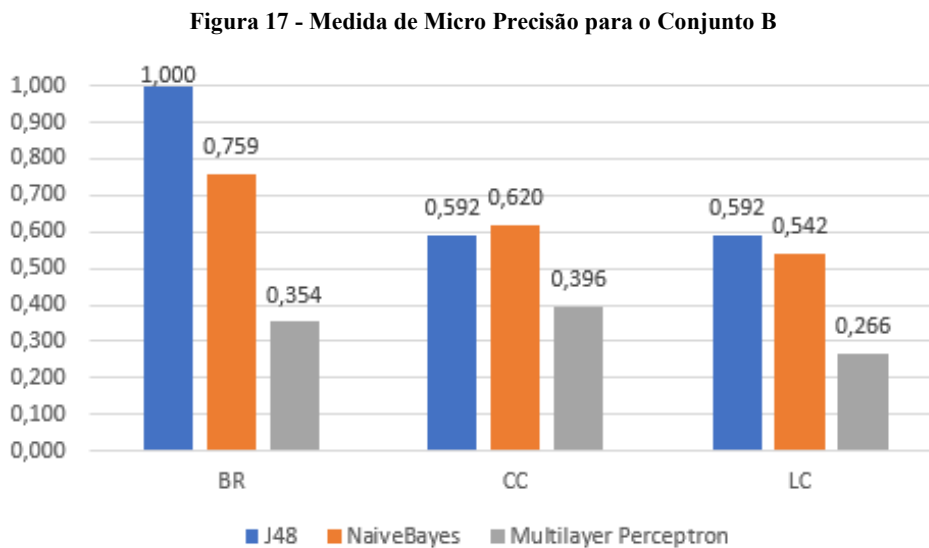


Fonte: Autor (2022)

Como podemos observar pelo gráfico da Acurácia, dentre todos os algoritmos testados o que apresentou melhores resultados foi o CC com J48. Ao utilizar o Conjunto A sua acurácia se mostrou melhor em comparação ao Conjunto B, que possui mais rótulos.



Fonte: Autor (2022)

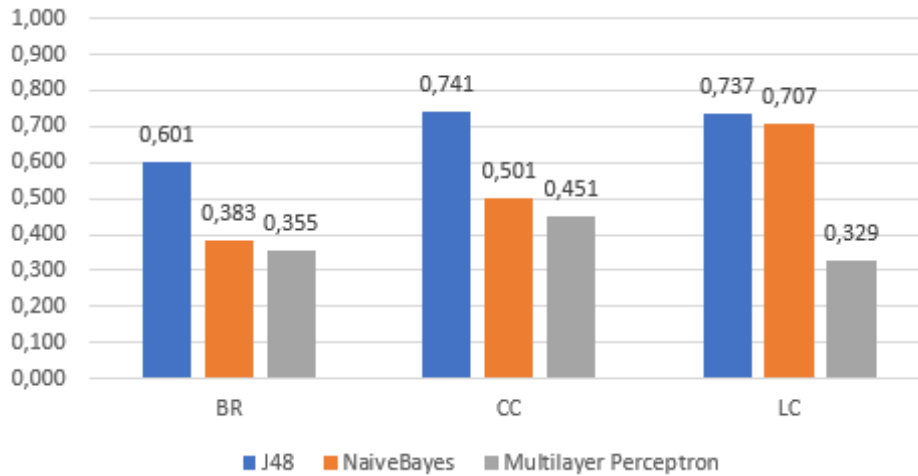


Fonte: Autor (2022)

Nesta medida avaliativa, pode-se notar a similaridade de seus resultados, mesmo um dos modelos tendo rótulos a menos. Esta medida visa mostrar quantas classificações estão corretas e pode ser usada para quando existem falsos positivos prejudiciais na recomendação.

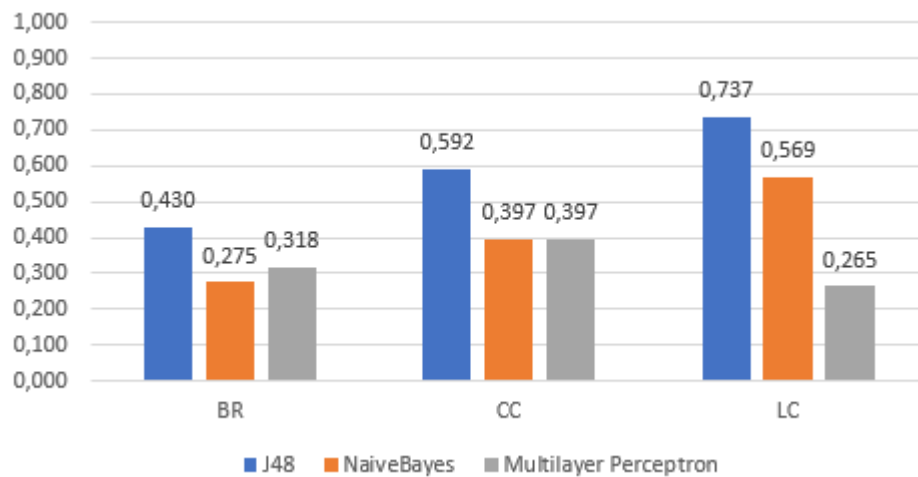


**Figura 18 - Medida de Micro *Recall* para o Conjunto A**



Fonte: Autor (2022)

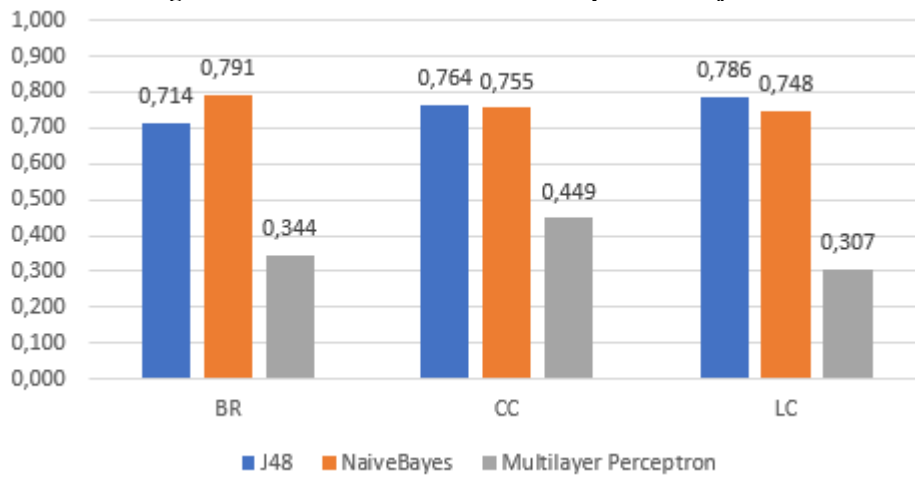
**Figura 19- Medida de Micro *Recall* para o Conjunto B**



Fonte: Autor (2022)

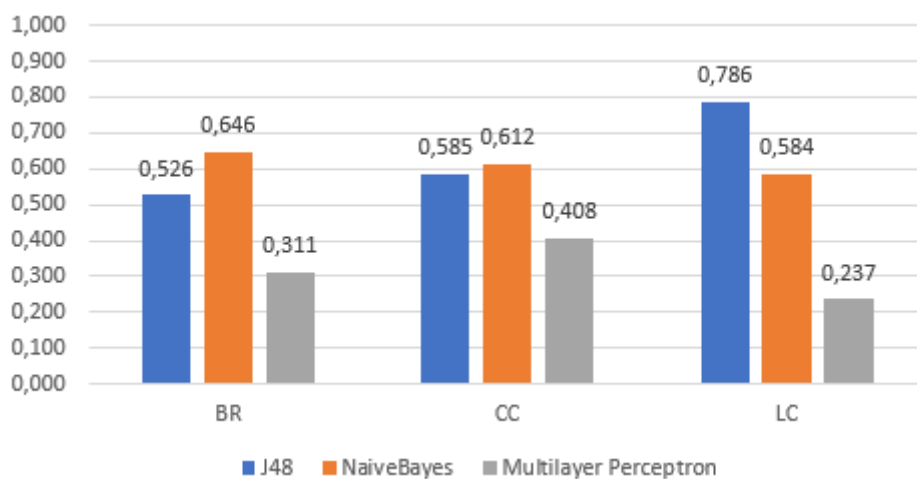
No caso da medida de Micro *Recall*, ela é usada em casos em que os falsos negativos são considerados mais perigosos que os falsos positivos, ou seja, altos índices de micro recall pode ser benéfico para a recomendação.

**Figura 20 - Medida de Macro Precisão para o Conjunto A**



Fonte: Autor (2022)

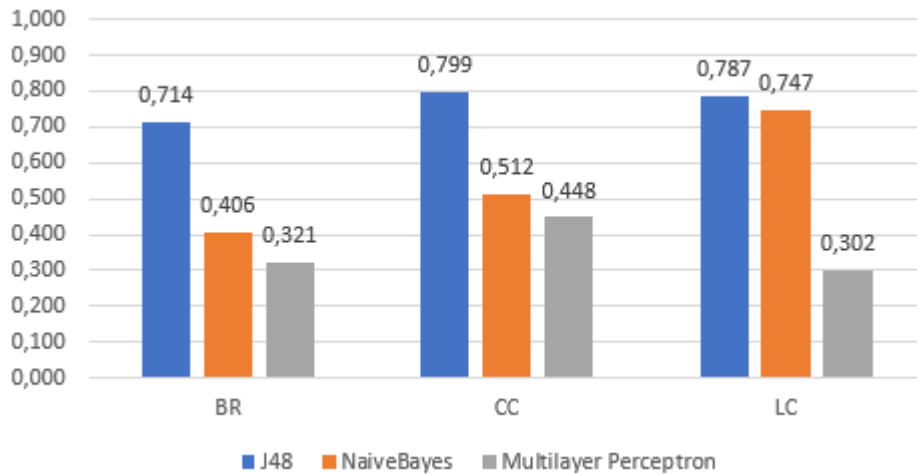
**Figura 21- Medida de Macro Precisão para o Conjunto B**



Fonte: Autor (2022)

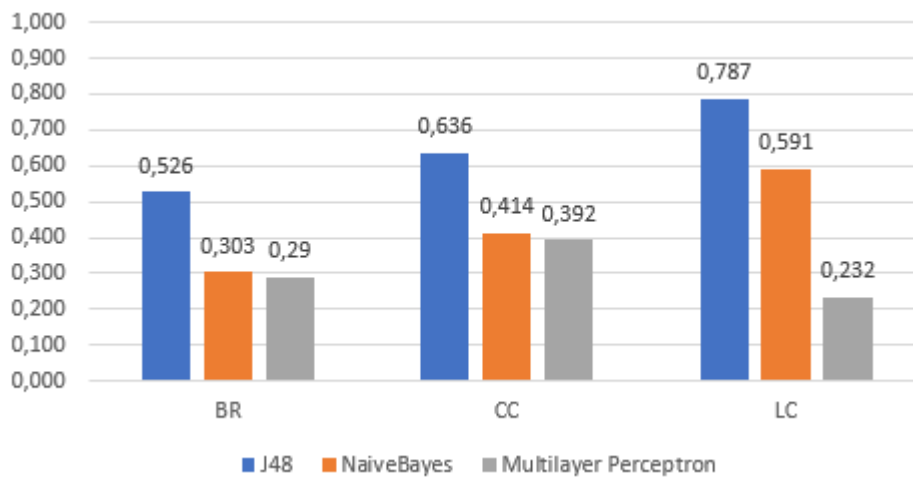
A medida de macro precisão são influenciadas por rótulos mais raros, pois é calculada com cada rotulo individualmente e depois é realizada uma média entre todos os rótulos, fazendo com que pesos iguais sejam aplicados independente se o rótulo é mais frequente ou não. Quanto mais próximo de 1,000 melhor a macro precisão.

**Figura 22 - Medida de Macro Recall para o Conjunto A**



Fonte: Autor (2022)

**Figura 23 - Medida de Macro Recall para o Conjunto B**

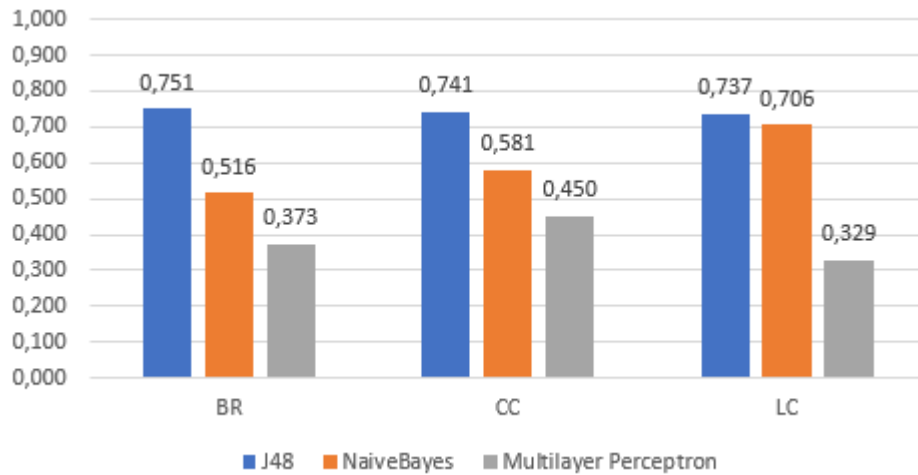


Fonte: Autor (2022)

Da mesma forma que acontece nas medidas de macro precisão, as medidas de Macro *Recall* também são influenciadas por rótulos raros, pois seu cálculo é muito parecido com a macro precisão. Esta medida responde a seguinte pergunta: Realmente este elemento é da classe  $x$ ? quão frequente ele é classificado como  $x$ ?

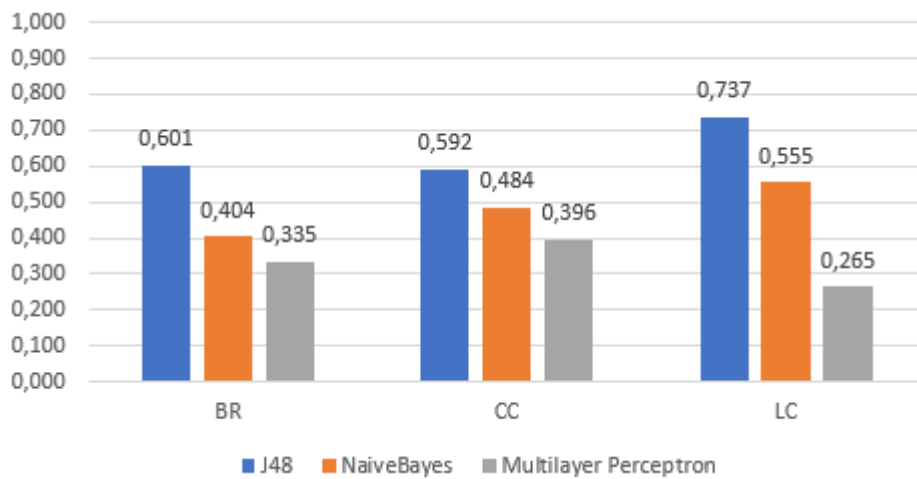
Assim como a medida anterior, quanto mais próximo de 1,000 melhor sua recomendação.

**Figura 24 - Medida de F1(Micro Averaged) do Conjunto A**



Fonte: Autor (2022)

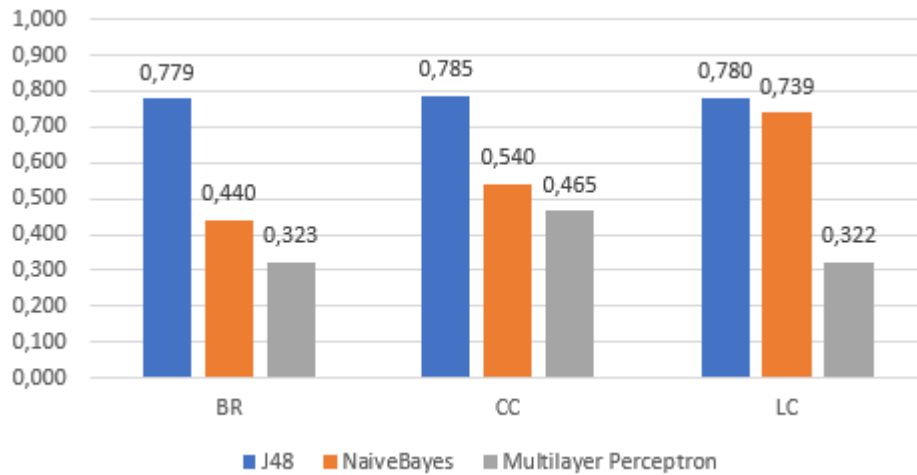
**Figura 25- Medida de F1(Micro Averaged) do Conjunto B**



Fonte: Autor (2022)

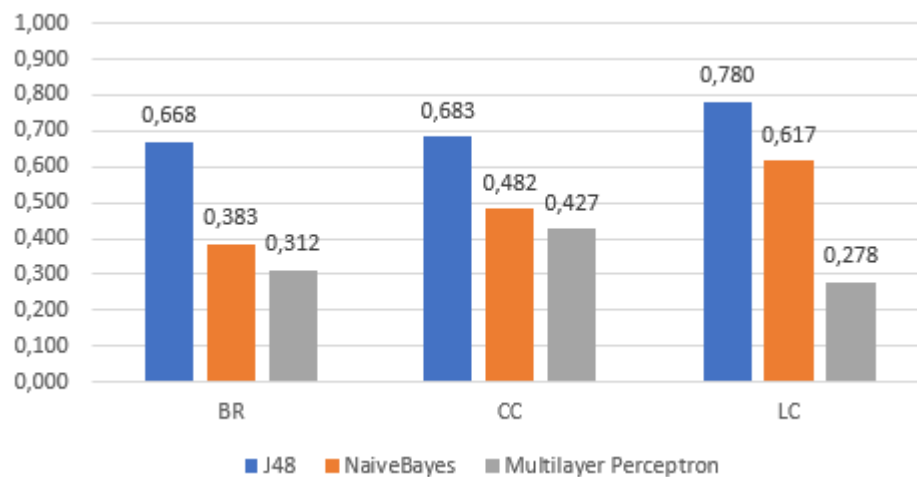
As medidas F1 levam em consideração os acertos do rótulo para cada nova combinação feita. Nessa medida quanto maior o valor, maior o acerto. É uma medida que também pode indicar se a precisão ou o recall está baixo.

**Figura 26 - Medida de F1 (*Macro Avarage by Exemple*) do Conjunto A**



Fonte: Autor (2022)

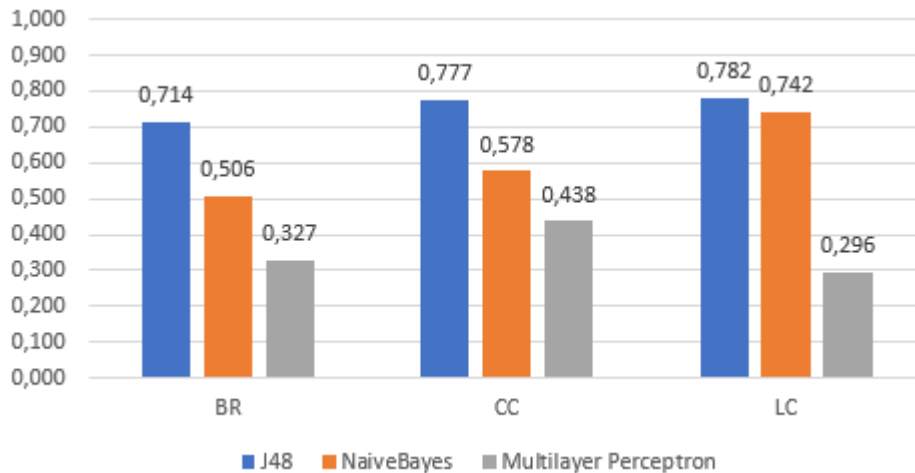
**Figura 27 - Medida de F1 (*Macro Avarage by Exemple*) do Conjunto B**



Fonte: Autor (2022)

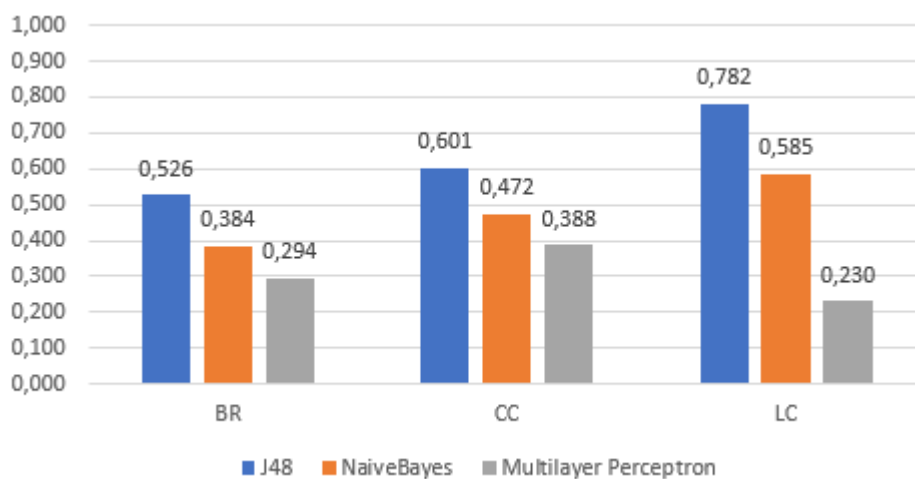
Assim como a medida anterior, aqui também é medido o F1 dos rótulos, mas observando a média macro dos dados dispostos. Quanto maior essa medida, melhor.

**Figura 28 - Medida de F1 (*Macro Avarage by Label*) do Conjunto A**



Fonte: Autor (2022)

**Figura 29 - Medida de F1 (*Macro Avarage by Label*) do Conjunto B**



Fonte: Autor (2022)

Avaliando as métricas é possível observar que o Conjunto A, com menos rótulos foi melhor que o Conjunto B, com mais rótulos, mas para confirmar essa diferença foi feito um teste estatístico não-paramétrico de Wilcoxon pareado somente na métrica acurácia, visto que ela demonstra uma visão geral do modelo de classificação obtido.

No teste foram consideradas as hipóteses com confiança de 5%:

- $H_0$ : Conjunto A e Conjunto B produzem os mesmos resultados
- $H_1$ : o Conjunto A produz resultados melhores que o Conjunto B

Com um p-valor de 0.001953, é possível dizer que, estatisticamente, em relação a acurácia, o Conjunto A produz resultados melhores utilizando menos rótulos e, assim como

demonstrado nas métricas é indicado como um melhor modelo para ser utilizado em um sistema de recomendação.

## 5. CONCLUSÃO

Ainda em março de 2022, o Brasil atingiu recorde de 77,5% de endividamento, é o maior índice nos últimos 12 anos de pesquisa, tudo isso leva a crer que a situação no que se diz respeito ao endividamento ainda vai demorar muito a se estabilizar, principalmente após uma pandemia que levou dois anos para dar um sinal de melhora.

Apesar de toda incerteza, uma das maiores preocupações do trabalhador brasileiro é estar em dia com suas contas. Com isso empresas como a Serasa buscam anualmente realizar feirões que tendem a incentivar o pagamento das mesmas, com descontos chegando na casa dos 90%.

Observando as medidas apresentadas para avaliação dos modelos de recomendação, quando se observa a acurácia de recomendação de ações para os atendentes, e se compara com o último dado proveniente do Serasa, onde, em média a taxa de recuperação de crédito é de 46,5% dos débitos de empresas após um prazo de até 60 dias de negativação (SERASA, 2022), ou seja, ainda dentro do período exigido pelo método de *collection score*, é possível concluir que em um modelo de recomendação de ações onde a acurácia é em média 71% é possível realizar recomendações voltadas a cobrança que podem ser benéficas na hora de abordar um devedor e, aumentar a taxa de recuperação de crédito.

Estas ações podem ser trabalhadas em trabalhos futuros que envolvam a criação desse sistema, além de beneficiar empresas que buscam recuperar dívidas em atraso ainda em prazo curto, como é o método *collection score*, ainda pode auxiliar pessoas que buscam quitar suas pendencias oferecendo descontos atrativos, formas de parcelamento, forma de pagamento entre outras ações que podem ser inclusas nesse sistema.



## REFERÊNCIAS

- AGNOLO, H. O. D, MOURA, G. W, CUNHA D. V. **Sistema para disponibilização e recomendação de eventos baseados em histórico e relacionamento do usuário.** 2021. Faculdade de Computação e Informática – Universidade Presbiteriana Mackenzie, São Paulo, 2021. Disponível em: <https://dspace.mackenzie.br/bitstream/handle/10899/27572/403-Artigo-HENRIQUE%20OLIVARES%20DELL%20AGNOLO.....pdf>. Acesso em: 29 nov. 2021
- ALIGER. **Entenda o aprendizado não supervisionado no Machine Learning.** 30 jul. 2019. Disponível em: <https://www.aliger.com.br/blog/machine-learning-entenda-o-que-aprendizado-nao-supervisionado/>. Acesso em: 25 nov. 2021
- BALDÁRRAGO, A. E. U. **Abordagem de Recomendação baseada em Conteúdo utilizando Ontologia Fuzzy de Domínio e Ontologia Crisp de Preferência do Usuário.** 2012. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de São Carlos, São Carlos, 2012. Disponível em: <https://repositorio.ufscar.br/bitstream/handle/ufscar/502/4477.pdf?sequence=1&isAllowed=y>. Acesso em: 27 nov. 2021
- BARBOSA, C. E. M. **Estudo de técnicas de Filtragem Híbrida em Sistemas de Recomendação de Produtos.** 2014. Monografia (Bacharelado em Ciência Computação) – Universidade Federal de Pernambuco, Recife, 2014. Disponível em: <https://www.cin.ufpe.br/~tg/2013-2/cemb.pdf>. Acesso em: 29 nov. 2021
- BARROS, P. **Aprendizagem de Máquina: Supervisionada e Não Supervisionada.** 7 abr. 2016. Disponível em: <https://medium.com/opensanca/aprendizagem-de-maquinasupervisionada-ou-n%C3%A3o-supervisionada-7d01f78cd80a>. Acesso em: 25 nov. 2021
- BONIN, M. **O que são sistemas de recomendação? Veja exemplos.** 6 set. 2018. Disponível em: <https://king.host/blog/2018/09/o-que-sao-sistemas-de-recomendação>. Acesso em: 27 nov. 2021
- CARVALHO, L. A. M. C, MACEDO, H. T. Introdução aos Sistemas de Recomendação para Grupos. **Revista de Informática Teórica e aplicada**, v. 21, n. 1, 2014. Disponível em: <https://www.seer.ufrgs.br/rita/article/view/CARVALHO-RITA-VOL21-NR1>. Acesso em: 27 nov. 2021
- CERRI, R. **Técnicas de classificação hierárquica multirrótulo.** 2010. Dissertação (Mestrado em Ciências da Computação e Matemática Computacional) – Universidade de São Paulo – USP, São Carlos, 2010. Disponível em: <https://teses.usp.br/teses/disponiveis/55/55134/tde-06042010-151017/publico/dissertacaoRicardoCerri.pdf>. Acesso em: 12 jun. 2022
- CERQUEIRA, P. H. R. **Um estudo sobre reconhecimento de padrões: um aprendizado supervisionado com classificador bayesiano.** 2010. Dissertação (Mestrado em Ciências) – Escola superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo. Piracicaba, 2010. Disponível em: <https://www.teses.usp.br/teses/disponiveis/11/11134/tde->

11022011101732/publico/Pedro\_Henrique\_Ramos\_Cerquino.pdf. Acesso em: 25 nov.2021

D'ADDIO, R. M. **Filtragem baseada em conteúdo auxiliada por métodos de indexação colaborativa**. 2015. Dissertação (Mestrado em Ciência da Computação e Matemática) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2015. Disponível em: [https://www.teses.usp.br/teses/disponiveis/55/55134/tde26112015144141/publico/RafaelDadio\\_revisada.pdf](https://www.teses.usp.br/teses/disponiveis/55/55134/tde26112015144141/publico/RafaelDadio_revisada.pdf). Acesso em: 27 nov. 2021

DONATI, A. **Concepção e desenvolvimento de um sistema de recomendação para o varejo físico**. 2018. Monografia (Bacharelado em Engenharia de Controle de Automação) – Universidade Federal de Santa Catarina, Florianópolis, 2018. Disponível em: [https://repositorio.ufsc.br/bitstream/handle/123456789/200001/PFC%20Andrei%20Donati\\_2018-2.pdf?sequence=1&isAllowed=y](https://repositorio.ufsc.br/bitstream/handle/123456789/200001/PFC%20Andrei%20Donati_2018-2.pdf?sequence=1&isAllowed=y). Acesso em: 30 nov. 2021

FORTI, M. **Técnicas de Machine Learning aplicadas na recuperação de crédito do mercado brasileiro**. 2018. Dissertação (Mestrado em Economia) - Faculdade de Economia, Fundação Getúlio Vargas. São Paulo, 2018. Disponível em: [http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/24653/Melissa\\_Forti\\_dissertacao.pdf?sequence=1&isAllowed=y](http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/24653/Melissa_Forti_dissertacao.pdf?sequence=1&isAllowed=y). Acesso em: 17 out. 2021.

GOMES, P. C. T. G. **Introdução ao aprendizado de máquina**. 14 mar. 2019. Disponível em: <https://www.datageeks.com.br/aprendizado-de-maquina/>. Acesso em: 24 nov. 2021.

GONÇALVES, E. C. Introdução à classificação Multirrótulo. **Minicurso de ERSI-RJ 2018 – V Escola Regional de Sistemas de Informação do Rio de Janeiro**, Rio de Janeiro, Cap. 3, dez. 2018. DOI: <https://doi.org/10.5753/sbc.7>. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/7/11/36-1>. Acesso em: 28 abril 2022.

G1. **Desemprego cai para 10,5% em abril e atinge 11,3 milhões, diz IBGE**. Disponível em: <https://g1.globo.com/economia/noticia/2022/05/31/desemprego-fica-em-105percent-em-abril.ghtml>. Acesso em: 22 jun. 2022

LIMA, I. R. R. C. **Recomendação de artigos científico utilizando filtragens colaborativa e híbrida**. 2012. Monografia (Bacharelado em Engenharia de Computação) – Universidade Estadual de Feira de Santana, Feira de Santana, 2012. Disponível em: <http://www.ecomp.uefs.br/data/tcc/RECOMENDA%C3%87%C3%83O%20DE%20ARTIGOS%20CIENT%3%8DFICOS%20UTILIZANDO%20FILTRAGENS%20COLABORATIVA%20E%20H%3%8DBRIDA.pdf>. Acesso em: 29 nov. 2021

LUDERMIR, T. B. **Inteligência artificial e aprendizado de máquina: estado atual e tendências**. 19 abr. 2021. Disponível em: <https://www.scielo.br/j/ea/a/wXBdv8yHBV9xHz8qG5RCgZd/>. Acesso em: 24 nov. 2021.

MACHADO, A. NICOLAY, S. **Estratégias de análise e concessão de crédito utilizadas por empresas industriais calçadistas: O caso do valo do Paranhana/RS**. Revista de Administração de Empresas Eletrônicas – RAEE, n. 7, 2017. Disponível em:

<https://seer.faccat.br/index.php/administracao/article/view/677>. Acesso em: 07 de jun. 2022

MACIEL, C. **Mapa da inadimplência aponta mais de 62 milhões de endividados**. AGÊNCIA BRASIL. 2 jul. 2021. Disponível em: <https://agenciabrasil.ebc.com.br/economia/noticia/2021-07/mapa-da-inadimpenciaaponta-mais-de-62-milhoes-de-endividados>. Acesso em: 03 out. 2021

MEDEIROS, I. R. G. **Estudo sobre Sistemas de Recomendação Colaborativos**. 2013. Monografia (Bacharelado em Ciência da Computação) - Universidade Federal de Pernambuco, Recife, 2013. Disponível em: <https://www.cin.ufpe.br/~tg/2012-2/irgm.pdf>. Acesso em: 15 out. 2021

MOTTA, C. L. R. et al. Sistemas de recomendação. **Pimentel, M.; Fuks, H. “Sistemas colaborativos”**. Rio de Janeiro: Elsevier, 2011. Disponível em: [https://www.researchgate.net/profile/JoniceOliveira/publication/328228374\\_Sistemas\\_de\\_Recomendacao/links/5bbf9ec5a6fdcc2c91f6ac1a/Sistemas-de-Recomendacao.pdf](https://www.researchgate.net/profile/JoniceOliveira/publication/328228374_Sistemas_de_Recomendacao/links/5bbf9ec5a6fdcc2c91f6ac1a/Sistemas-de-Recomendacao.pdf). Acesso em: 27 nov. 2021

NOGUEIRA, E. A. **Uma abordagem baseada em filtragem colaborativa integrada a mapas de saliência para a recomendação de imagens**. 2015. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Uberlândia, Uberlândia, 2015. Disponível em: <https://repositorio.ufu.br/bitstream/123456789/17682/1/AbordagemFiltragemColaborativa.pdf>. Acesso em: 30 nov. 2021

NEVES, M. DEUS, C. **Concessão de Crédito**. 2015. XII Simpósio de Ciências Aplicadas da FAIT – Faculdade de Ciências Sociais e Agrárias de Itapeva, Itapeva, 2017. Disponível em: [http://fait.revista.inf.br/imagens\\_arquivos/arquivos\\_destaque/8bBzdOEsBU2RhO7\\_2017-1-17-19-12-3.pdf](http://fait.revista.inf.br/imagens_arquivos/arquivos_destaque/8bBzdOEsBU2RhO7_2017-1-17-19-12-3.pdf). Acesso em: 07 de jun. 2022

O GLOBO. **Endividamento e inadimplência das famílias batem novo recorde em abril**. Disponível em: <https://g1.globo.com/economia/noticia/2022/05/02/endividamento-e-inadimplencia-das-familias-batem-novo-recorde-em-abril.ghtml>. Acesso em: 22 jun. 2022.

PAULA, L. F. de, JÚNIOR, A. J. A, **Comportamento dos bancos e ciclo de crédito no Brasil em 2003 – 2016: Uma análise pós Keynesiana da preferência pela liquidez**. 7 ago. 2020. Disponível em: <https://www.scielo.br/j/rec/a/4gP8qkcXWtJycnDY8P7t76t/?lang=pt>. Acesso em: 21 nov. 2021

PELLUCCI, P.R.S. *et al.* **Utilização de técnicas de Aprendizado de Máquina no reconhecimento de entidades nomeadas no Português**. Centro Universitário de Belo Horizonte, Belo Horizonte – MG, 2011, Disponível em: <https://core.ac.uk/download/pdf/235250996.pdf>. Acesso em 12 de jun. 2022

PINHEIRO, H. **Serie AI/ML Cold Recommendation**. 6 mar. 2019. Disponível em: <https://medium.com/codengage/serie-ai-ml-cold-recommendation-59d514512300>. Acesso em: 04 dez. 2021

RICH, E. KNIGHT, K. “**Artificial Intelligence**”. 2. ed. McGraw-Hill Education. 1991

ROLIM V.B. *et al.* Um estudo sobre Sistemas de Recomendação de Recursos Educacionais.2018. **CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO**. Disponível em: [https://www.researchgate.net/publication/320993342\\_Um\\_Estudo\\_Sobre\\_Sistemas\\_de\\_Recomendacao\\_de\\_Recursos\\_Educacionais](https://www.researchgate.net/publication/320993342_Um_Estudo_Sobre_Sistemas_de_Recomendacao_de_Recursos_Educacionais). Acesso em: 30 nov. 2021

SANTOS, L. H. A; ROSSI, R. G. **Aprendizado Multirrótulo para Classificação Automática de Textos** – Universidade Federal do Mato Grosso do Sul, Mato Grosso do Sul, 2020. Disponível em: [https://cptl.ufms.br/files/2020/02/Aprendizado\\_Multirr\\_tulo\\_para\\_a\\_Classifica\\_\\_o\\_Autom\\_tica\\_de\\_Textos1.pdf](https://cptl.ufms.br/files/2020/02/Aprendizado_Multirr_tulo_para_a_Classifica__o_Autom_tica_de_Textos1.pdf). Acesso em 28 abr. 2022.

SERASA, **Mapa da inadimplência e renegociação de dívidas no Brasil. Levantamento mensal da Serasa sobre a relação dos brasileiros com as dívidas**. Mar 2022. Disponível em: <https://www.serasa.com.br/assets/cms/2022/Mapa-da-inadimplencia-MARCO.pdf>. Acesso em: 07 jun. 2022

TAKAHASHI, M. **Estudo comparativo de algoritmos de recomendação**. – Universidade de São Paulo, São Paulo, fev. 2015. Disponível em: [https://bcc.ime.usp.br/tccs/2014/marcost/monografia\\_final.pdf](https://bcc.ime.usp.br/tccs/2014/marcost/monografia_final.pdf). Acesso em: 18 jun. 2022

UOL. **Desemprego no país mantém recorde de 14,7% e atinge 14,8 milhões, diz IBGE**. 30 jun. 2021. Disponível em: <https://economia.uol.com.br/empregos-ecarreiras/noticias/redacao/2021/06/30/desemprego-pnad-ibge.htm>. Acesso em: 03 out. 2021

SILVA, C. **Machine Learning e sua relação com a Ciência da Computação**. 2020. Disponível em: <https://inscricao.ucb.catolica.edu.br/blog/machine-learning>. Acesso em: 24 nov. 2021.

SILVA, R. G. N. **Sistema de Recomendação baseado em conteúdo textual: avaliação e comparação**. Universidade Federal da Bahia, Salvador – BA. 2014. Disponível em: [https://repositorio.ufba.br/bitstream/ri/19281/1/dissertacao\\_mestrado\\_ciencia\\_computacao\\_rafael\\_glauber.pdf](https://repositorio.ufba.br/bitstream/ri/19281/1/dissertacao_mestrado_ciencia_computacao_rafael_glauber.pdf). Acesso em: 18 jun. 2022

SOUSA, J. ASSERTIVA, **Collection score: como funciona e quais as vantagens para o seu negócio?** 10 out. 2018. Disponível em: <https://blog.assertivasolucoes.com.br/collection-score/>. Acesso em: 12 nov. 2021

ZUÑIGA, N. ASSERTIVA, **Ciclo de Crédito: Como mapear a jornada do seu cliente**. 20 dez. 2019, Disponível em: <https://blog.assertivasolucoes.com.br/ciclo-de-creditocomo-mapear-a-jornada-do-seu-cliente/>. Acesso em: 12 nov. 2021