

# Rotule-me! Uma experiência de engenharia de requisitos para um sistema de rotulagem

Letícia Carvalho Passos<sup>1,2</sup>[0000-0002-8308-032X], Lucas Lima Florêncio Viana<sup>1,2</sup>[0000-0002-8154-0062], Edson César de Oliveira<sup>2,3</sup>[0000-0001-9168-4388] e Tayana Uchoa Conte<sup>1,2</sup>[0000-0001-6436-3773]

<sup>1</sup> Universidade Federal do Amazonas – UFAM

Instituto de Computação - IComp

<sup>2</sup> USES Research Group

<sup>3</sup> Secretaria do Estado da Fazenda do Estado do Amazonas - SEFAZ-AM  
{leticia.passos, lucas, tayana}@icomp.ufam.edu.br,  
edson.cesar@sefaz.am.gov.br

**Abstract.** Machine Learning (ML) tem sido cada vez mais aplicado em diversos domínios do conhecimento humano. Em projetos de ML supervisionado, é necessário que haja uma base de dados para treinamento do modelo de ML. Sistemas de rotulagem de dados são frequentemente utilizados para apoiar o desenvolvimento desse tipo de ML quando a gestão dessa base de dados é crítica para o sucesso do projeto. Contudo, a elicitación de requisitos desses sistemas apresenta desafios relacionados à interação dos usuários com grande volume de dados. Há a necessidade de entender os requisitos específicos desses sistemas para elaborar um software de qualidade. Este trabalho relata uma experiência de engenharia de requisitos em um sistema de rotulagem de dados para ML. Esse sistema tem como objetivo atuar na inspeção de rotulagem de milhões de produtos para fiscalização de impostos pelas Secretarias da Fazenda dos Estados Federativos do Brasil. A partir dos requisitos iniciais, foi realizado um processo evolutivo de prototipagem de interfaces. O processo incluiu três iterações, envolvendo o gerente de produto e a equipe de desenvolvimento. Os protótipos auxiliaram a refinar os requisitos a partir da análise de cada iteração, além de possibilitar a avaliação dos requisitos atendidos.

**Keywords:** requirements engineering · visual analytics · machine learning · data labeling systems.

## 1 Introdução

A demanda por sistemas de Machine Learning (ML) tem aumentado nos últimos tempos. Os sistemas ML são aplicados em diversos domínios, desde medicina [16] até o ensino de programação [11]. A utilização de ML somente é viável se existir uma base de dados de qualidade. Isso significa que a base deve ter: uma quantidade de dados balanceada, as classificações da base devem estar corretas e

deve existir uma grande quantidade de dados para que o modelo de conhecimento seja treinado corretamente [2].

Visando obter uma grande quantidade de dados, a rotulagem de dados pode utilizar recursos de outros campos de pesquisa como o Visual Analytics (VA) para trazer uma forma fácil e eficaz na tarefa de rotulagem de dados em grandes quantidades. O VA tem como objetivo viabilizar aos usuários a interpretação e exploração em grandes volumes de dados, muitas vezes proporcionando metáforas visuais dos dados para haver uma melhor interação do usuário [18,12].

Um dos desafios encontrados em sistemas de rotulagem, que utilizam ou não o VA, é a forma de interação entre o usuário e os dados [8], de tal modo que não seja cansativo, a ponto do usuário desistir do sistema [10]. Portanto, um dos objetivos dos sistemas de rotulagem é: “A interface deve apoiar a análise humana diante de grandes volumes de dados”.

Diante dos benefícios dos sistemas de rotulagem para apoiar a análise humana em grandes volumes de dados, este trabalho relata uma pesquisa feita no âmbito das Secretarias da Fazenda dos Estados Federativos do Brasil, que decidiram investigar como melhorar o processo de fiscalização de impostos com o auxílio do ML. Assim, fez-se necessário iniciar um projeto de um sistema de rotulagem de dados para atuar na inspeção das categorias dos produtos comercializados.

As Secretarias da Fazenda desde 2005 já processaram mais de 26.614 bilhões de Notas Fiscais Eletrônicas (NF-e’s) e cada NF-e possui até 200 produtos [21]. Por conta desse grande volume de dados, é humanamente impossível o usuário classificar cada um desses produtos. Além disso, a equipe não possui experiência prévia no desenvolvimento de sistemas ML, como também em sistemas de rotulagem de dados.

Finalmente, dado o contexto do problema descrito anteriormente, o objetivo desta pesquisa é investigar a elicitação de requisitos para este tipo de sistema com a finalidade de elaborar um software de qualidade. Os passos realizados para alcançar o objetivo da pesquisa foram: busca exploratória em revisões da literatura existentes com foco em identificar outros sistemas de rotulagem de dados [5,3,15], ciclo de 3 iterações com a equipe de desenvolvimento e uma análise dos requisitos de sistema atendidos.

As seções deste trabalho são divididas em: fundamentação teórica (Seção 2), Sistema de Rotulagem de Dados: Experiência nas Secretarias da Fazenda (Seção 3), Soluções Propostas (Seção 4), Discussão (Seção 5) e por fim, Conclusão (Seção 6).

## 2 Fundamentação Teórica

Nesta seção são explanados os conceitos fundamentais para o entendimento deste trabalho. São abordados os conceitos de visual analytics, machine learning e sistemas de rotulagem.

## 2.1 Visual Analytics

Cada vez mais é aproveitada a grande quantidade de dados que são capturados e criados pelos sistemas de informação da atualidade. Estes dados são investigados para se adquirir conhecimentos dentro de um determinado domínio, porém a investigação destes em grandes volumes é limitada para as pessoas pela dificuldade de interpretá-los [4].

Muitas iniciativas vêm sendo realizadas em torno de ajudar as pessoas na interpretação de dados em grande quantidade [18,1]. Essas iniciativas envolvem a criação ou uso de um sistema com técnicas que ajudam as pessoas na visualização e interpretação dos dados. A área que investiga estes sistemas e técnicas é o Visual Analytics (VA) e este pode ser aplicado em diversas áreas de conhecimento [16,12].

Há um fator que pode limitar a utilização do VA para a análise dos dados que é o próprio tipo de dado. Por exemplo, quando o tipo do dado é imagem, os recursos visuais que possibilitam o usuário interagir são muito intuitivos devido à própria natureza do dado [12]. Recursos como editar a imagem, recortar, inserir elementos, entre outros são de uso comum na área de manipulação e edição de imagens e servem como base para esses sistemas de VA.

Quando o tipo de dado é textual, os recursos desses sistemas devem auxiliar os usuários na interpretação dos textos. Porém, a elaboração dos recursos com textos não é trivial [12], pois a maioria dos textos não está disponibilizada de forma estruturada, ou seja, de fácil compreensão. Apesar dos desafios encontrados devido a cada tipo de dado, os sistemas de VA possuem muitos benefícios na análise dos dados, devido à redução do tempo, aumento da precisão e diminuição de custos da análise [18].

## 2.2 Machine Learning

ML é uma subárea da Inteligência Artificial que compõe algoritmos capazes de melhorar seu desempenho com a experiência. Um algoritmo neste domínio é capaz de aprender a partir de dados, capturando padrões e efetuando inferências [14].

Para que ocorra o aprendizado, os algoritmos de ML passam por um processo de aquisição da experiência, chamado de treinamento. Durante essa fase, o algoritmo busca uma função, que mapeie os dados de entrada e gere um valor na saída. Estes modelos podem ser agrupados em diferentes categorias quanto ao tipo de aprendizagem e a saída desejada para o algoritmo [7].

Quanto ao tipo de aprendizagem, as tarefas de ML podem ser agrupadas em aprendizagem supervisionada ou não-supervisionada. Na aprendizagem supervisionada, o algoritmo deve aprender a inferir a partir de exemplos rotulados por especialistas ou coletados na vida real. Por outro lado, na aprendizagem não-supervisionada, o algoritmo deve inferir padrões e estruturas a partir de dados não rotulados [7].

Um dos principais fatores para um modelo de ML ter bons resultados é possuir uma base de treinamento em grande quantidade e com qualidade nos

dados. Caso contrário, o modelo não aprenderá a inferir corretamente, mesmo sendo desenvolvido com os melhores algoritmos [7].

### 2.3 Sistemas de rotulagem

Os algoritmos de ML supervisionado necessitam de muitos dados rotulados para aprender. Isso quer dizer que os dados precisam estar previamente classificados. Se esta classificação depender de pessoas, é necessário apresentar estes dados da melhor forma possível, por isso muitos métodos de VA foram propostos para a rotulagem interativa. Reduzir o esforço de rotulagem é um dos principais objetivos dos sistemas de rotulagem [23].

Sistemas de rotulagem apresentam muitos dados para serem rotulados para os usuários. Dependendo do domínio dos dados, estes usuários podem ser especialistas ou não, como é o caso dos sistemas de rotulagem *crowdsorce*. Estes sistemas contam com milhares de usuários para realizar a tarefa de rotular os dados. Então, em seguida, a validação destas rotulagens pode ser feita por especialistas ou usuários experientes [23].

## 3 Sistema de Rotulagem de Dados: Experiência nas Secretarias da Fazenda

Esta seção descreve o problema enfrentado pelas Secretarias da Fazenda dos Estados de todo o Brasil, que levou ao desenvolvimento do sistema de rotulagem de dados. Além disso, é detalhada a elicitação de requisitos deste sistema, o que motivou esta pesquisa.

As Secretarias da Fazenda já processaram, desde outubro de 2005, sob o projeto Nota Fiscal Eletrônica (NF-e), mais de 26.614 bilhões de NF-e's [21]. As NF-e's registram as operações comerciais realizadas nos estabelecimentos e indústrias, podendo, cada nota, conter até 200 descrições dos produtos comercializados.

A partir dos dados constantes nas NF-e's, os governos estaduais realizam o cálculo do Imposto sobre Circulação de Mercadorias e Serviços (ICMS), que, por fim, financiará todas as políticas governamentais, desde a educação até a saúde pública. Caso a identificação do produto descrito na NF-e seja incorreta, haverá injustiça fiscal, incorrendo na cobrança a maior do ICMS, prejudicando o contribuinte, ou a cobrança a menor, prejudicando o governo estadual [9].

Uma das variáveis do cálculo do ICMS é o código NCM ou Nomenclatura Comum do Mercosul [19]. Cada produto descrito nas NF-e's é classificado em um código NCM pelos contribuintes. Porém, é comum que existam muitos erros nestas classificações, seja pela dificuldade em classificar um produto em uma das 10.200 possibilidades de NCM, ou seja pela intenção de sonegação fiscal [19].

Tendo em vista as milhões de descrições de produtos em NF-e's recebidas pelo Fisco Estadual, é necessário verificar se a cada uma dessas descrições foi atribuído o código NCM correto pelos contribuintes. Entretanto, o grande volume de NF-e's recebidas pelas secretarias diariamente, torna humanamente impossível a tarefa de validar item a item os milhares de produtos comercializados.

A solução escolhida para resolver este problema foi desenvolver um modelo de ML para classificar automaticamente cada produto. Porém, é necessário, anteriormente, ter uma boa base de dados para treinamento [2]. Isto significa dizer que esta base precisa possuir uma grande quantidade de dados distintos que representem cada classe corretamente.

Então, para obter esta base para treinamento é necessário construir um sistema de rotulagem de dados onde especialistas avaliam a classificação de milhões de produtos de acordo com o código NCM correto [20]. Entretanto, esta tarefa de rotulagem dos dados demanda muito tempo dos especialistas, o que pode causar desconforto e desmotivação ao usar o sistema [6].

O sistema de rotulagem deste problema em específico considera que seus usuários são especialistas em classificar títulos de produtos para um determinado código NCM. Devido ao grande volume de dados a serem apresentados e a especialidade dos usuários, é necessário, portanto, propor soluções de interface para entender como melhor apresentar estes dados e como classificar cada título corretamente com menos esforço.

Para propor estas soluções, foram realizadas interações entre o gerente de produto e a equipe de pesquisa deste trabalho para definir quais eram os requisitos iniciais deste sistema. Estas interações tiveram como base um processo de elicitação baseado em prototipagem [13]. Por fim, na Tabela 1, são apresentados os requisitos principais iniciais para este sistema. Os requisitos foram elaborados considerando o objetivo do sistema de rotulagem.

**Tabela 1.** Requisitos principais iniciais do sistema de rotulagem.

| ID  | Requisitos Iniciais                                                                                                                                                                                                        |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| RI1 | O sistema deve possuir elementos na interface para acelerar o processo de rotulagem (por exemplo, palavras destacadas para facilitar a leitura dos dados ou visualização de múltiplas instâncias para rotulagem em grupo.) |
| RI2 | O sistema deve apresentar os produtos e as suas respectivas categorias para a rotulagem                                                                                                                                    |
| RI3 | O sistema deve fornecer informações complementares para auxiliar o usuário em tomadas de decisão quanto a tarefa de rotulagem                                                                                              |

Estes requisitos principais iniciais foram utilizados como base para a construção dos protótipos de interface. Com as apresentações dos protótipos, os requisitos foram refinados, tornando-se mais concretos durante as iterações do processo.

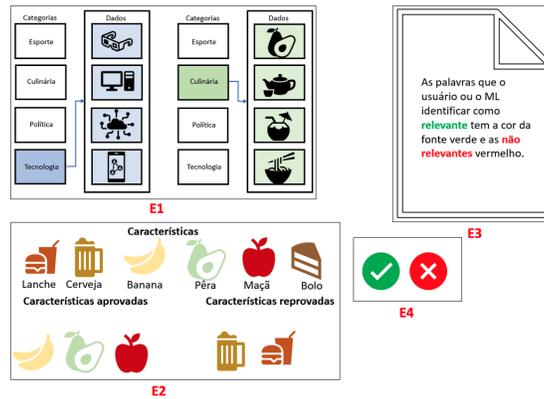
## 4 Soluções propostas

Esta seção apresenta os passos seguidos durante todo o processo de criação dos protótipos para o sistema de rotulagem do problema das Secretarias da Fazenda. Os passos realizados foram: pesquisa exploratória na literatura, processo de evolução dos protótipos criados e as soluções para sistemas de rotulagem.

#### 4.1 Pesquisa exploratória

Para resolver os problemas apresentados na Seção 3, foi primeiramente realizada uma pesquisa exploratória na literatura científica com o objetivo de analisar estudos sobre sistemas de rotulagem e seus elementos de interface. Foram analisados: o trabalho de Dudley et al. [5] e as RSL de Amershi et al. [3] e Nadj et al. [15].

O resultado da pesquisa exploratória permitiu identificar técnicas de VA aplicadas nos sistemas de rotulagem, através de elementos de interface que serviram como base para a criação dos protótipos do sistema de rotulagem proposto neste trabalho. Conforme a Figura 1, estes elementos foram identificados como: Listas Cruzadas (E1) [11], *Feedback* de Atributos Positivos e Negativos (E2) [3], Palavras Relevantes em Cores (E3) [22] e Escolha Binária Sobre a Relevância do Dado (E4) [17].



**Figura 1.** Técnicas de VA aplicadas nos sistemas de rotulagem encontrados na literatura.

1. **E1.** As listas cruzadas se dividem em duas listas, a do lado esquerdo apresenta as categorias de dados e a do lado direito são apresentados os dados de cada categoria conforme o usuário clique em uma delas. O E1 auxilia na visualização de dados de cada categoria numa disposição paralela entre as listas. Assim, é possível visualizar rapidamente várias instâncias pertencentes a um determinado grupo.

Na Figura 1.E1 são apresentados dois exemplos de listas cruzadas. No primeiro exemplo, a categoria selecionada é tecnologia (cor azul), e como consequência são apresentados os dados desta categoria. No segundo exemplo, a categoria selecionada é culinária (cor verde) e assim como no exemplo anterior, são apresentados os dados desta categoria.

2. **E2.** O usuário fornece *feedback* identificando atributos positivos e negativos. Este elemento agrupa essas características em um grupo de inclusão

e outro de exclusão. Assim, o usuário pode verificar os dados e excluir as características que não representem sua respectiva classe. Por exemplo, se o objetivo do usuário é classificar textos que tenham apenas as palavras banana, pêra e maçã, mas também não queira textos com as palavras cerveja e lanche, o resultado desse elemento será apresentar apenas os textos que atendam as características aprovadas e que não tenham as características reprovadas, conforme a Figura 1.E2.

3. **E3.** Palavras relevantes em cores destacam palavras importantes em um texto. Este elemento auxilia na visualização rápida de termos em um texto através de cores que cativem a atenção do usuário. Assim, a interpretação do texto na totalidade é agilizada e traz a atenção do usuário para o assunto central do dado analisado. Na Figura 1.E3, as palavras identificadas como relevantes possuem a cor da fonte verde e as não relevantes da cor vermelha.
4. **E4.** Escolha sobre a relevância do dado é uma interação do usuário ao indicar se um dado é relevante ou não. Este elemento auxilia na coleta do *feedback* do usuário sobre um determinado dado com o toque de um único botão. Além disso, este elemento é indicado para problemas de classificação binária, ou seja, escolher se um dado pertence em uma dentre duas classes. Na Figura 1.E4, são utilizados o ícone “check” para indicar o texto relevante e “x” para indicar o texto como não relevante.

## 4.2 Protótipo Inicial

Conforme a Figura 2, inicialmente pensou-se na criação de um protótipo modelado para a rotulagem dos códigos NCM associados aos produtos, ou seja, a tarefa do usuário seria de selecionar o código NCM correto dentre os códigos sugeridos. Neste protótipo, foi utilizado uma adaptação do 1.E4 como solução para obter *feedback* do usuário sobre o rótulo do produto com apenas um toque.

Outros elementos de interface também foram utilizados. Um deles foi um gráfico de barras mostrando os códigos NCM associados aos produtos que tinham uma determinada palavra na sua descrição, como também o status do andamento da tarefa.

Em conjunto com o gerente de produto do projeto foram avaliados os Pontos Fortes (PF) e fracos do protótipo apresentado na Figura 2. Os pontos fortes são:

- **PF1: gráfico de barra de uma palavra representativa de um produto (Figura 2.B)** O gráfico auxilia o usuário na escolha do código NCM correto apresentando os códigos NCM mais associados a um determinado produto.
- **PF2: status de andamento (Figura 2.C)** Através do status de andamento, o usuário é informado quanto a quantidade de produtos categorizados por ele e por outros usuários.

Porém, um ponto fraco encontrado é com relação à estratégia de rotulagem utilizada. Um dos pontos levantados pelo gerente de produto é o fato do usuário precisar avaliar o código NCM de cada produto, porém as secretarias recebem

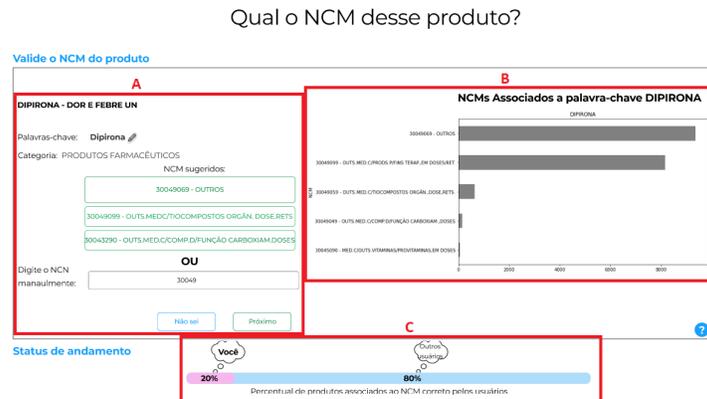


Figura 2. Protótipo inicial.

milhões de notas por dia o que torna inviável ao usuário avaliar todos os produtos. Outro ponto é que a avaliação do código NCM de cada produto pode provocar fadiga no usuário, o que pode afetar a qualidade da avaliação devido ao cansaço gerado durante a atividade de rotulagem.

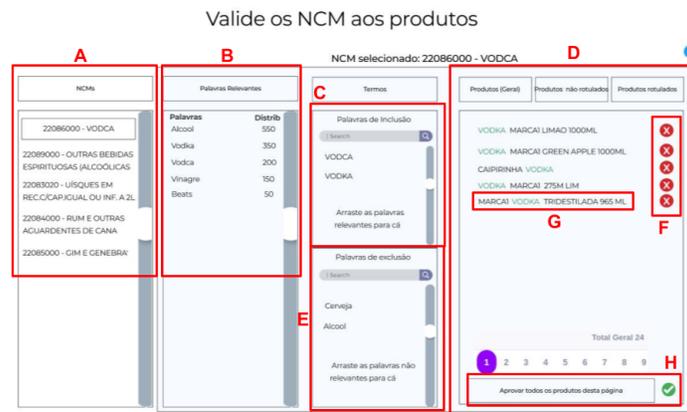
**Avaliação e lições aprendidas** Tendo em vista a inviabilidade da solução empregada no protótipo da Figura 2, foram discutidas outras estratégias para contornar este problema. Uma das estratégias que resolveria este problema é solicitar do usuário a avaliação de grupos de produtos associados a um código NCM, visto que estes códigos são um conjunto de grandeza menor (10.200) comparado com o número de produtos descritos nas NF-e's recebidas pelas secretarias por dia ( $\approx 1$  MI).

### 4.3 Segundo Protótipo

Para solucionar o problema encontrado pelo gerente de produto na proposta anterior, foi construído um novo protótipo analisando quais elementos podiam ser mais compatíveis. Assim, foram utilizadas todas as soluções E1, E2, E3 e E4 para este protótipo.

O protótipo apresentado na Figura 3 foi criado com foco na análise dos produtos por NCM. O usuário precisa primeiro selecionar um código NCM, conforme Figura 3.A. Em seguida, os componentes são carregados com as informações de todos os produtos associados ao código NCM selecionado, conforme o uso da solução E1.

Após a finalização do protótipo, novamente foi realizada uma validação do protótipo com o gerente de produto. Nesta validação, foram identificados os seguintes pontos fortes:



**Figura 3.** Segundo protótipo proposto.

- **PF3: Agrupamento e filtros de termos representativos (Figura 3.B)**  
 O filtro de termos apresenta todos os termos únicos dos produtos associados ao código NCM selecionado. Além disso, é apresentado a frequência que esses termos ocorrem, possibilitando ao usuário verificar quais os termos mais utilizados nos produtos. O uso da solução E2 permitiu que os termos fossem “arrastados” para os componentes de palavras de inclusão e exclusão conforme a Figura 3.C, assim estes termos refletem nos dados apresentados no componente D da Figura 3.
- **PF4: Destaque de termos aprovados e reprovados por cor (Figura 3.G)**  
 A inserção de termos no componente C e E da Figura 3 altera a cor da fonte na descrição do produto (Figura 3.G), sendo verde para indicar que a palavra é um termo de relevância para a categorização do produto e vermelho para indicar o termo como não relevante, conforme o uso da solução E3. Além disso, com o uso da solução E4, foi possível aprovar grupos de produtos ou reprovar um produto em específico.

Os pontos fracos identificados pelo gerente de produto no segundo protótipo foram:

- **Sobrecarga de informações no protótipo**  
 Apesar de todas as informações serem necessárias, existe uma sobrecarga de informações que pode confundir o usuário e dificultar a análise ao expor as informações de diversos códigos NCM junto com muitos títulos de produtos. Uma sugestão do gerente de produto foi a possibilidade de analisar os produtos pertencentes a um código NCM pré-selecionado anteriormente.
- **Impossibilidade de agrupar termos**  
 Há termos que combinados com outros devem ser classificados em conjunto como critérios de inclusão ou exclusão. Um exemplo seria de um código NCM que descreve os produtos de combustível do tipo álcool que, neste caso, possui o termo “álcool” como

o mais representativo, porém este em conjunto com o termo “gel” não deve ser considerado. Mesmo com a utilização da solução E4 para remover itens específicos do grupo, ainda se tornaria inviável remover item a item, caso existissem uma quantidade significativa desses produtos incorretos.

**Avaliação e lições aprendidas** O problema destacado pelo gerente de produto no protótipo inicial foi resolvido através da estratégia de apresentar os produtos agrupados por código NCM e assim possibilitando a avaliação de produto por grupo e não mais individualmente.

Além disso, neste protótipo foi utilizado o E2 como solução para a inclusão e exclusão de termos representativos, dado que alguns termos estão presentes em uma grande quantidade de produtos. Entretanto, foi levantada a necessidade de fazer a separação por grupos de termos e um design mais simples e limpo.

#### 4.4 Terceiro Protótipo

Dado os problemas relatados pelo gerente de produto no protótipo apresentado na Subseção 4.3, foi realizada uma nova iteração para incluir as soluções de elementos que atendam estes problemas. Novamente, foram utilizadas todas as soluções E1, E2, E3 e E4.

O novo protótipo foca no design mais simples através da avaliação de produtos de apenas um código NCM por vez, como pode ser visto na Figura 4. Além disso, neste protótipo é possível classificar grupos de termos, como também visualizar a frequência de palavras associadas a cada termo.



Figura 4. Terceiro protótipo proposto.

Após a finalização do protótipo, foi realizada uma nova validação de protótipo com o gerente de produto do projeto. Os seguintes pontos fortes foram identificados:

- **PF5: Design *clean* (Figura 4)** O novo protótipo aparenta conter menos informações por conta da exibição de produtos de apenas um NCM por vez. A visualização de produtos no componente C está associada a interação do usuário nos componentes B e E, conforme a utilização da solução E1.
- **PF6: Inclusão de grupos de termos para exclusão ou inclusão (Figuras 4.D e 4.F)** No componente F, é possível ver que o termo principal “MARCA2” em conjunto com “gin” foram reprovados pelo usuário. Assim, conforme utilização da solução E4, um termo pode ser aprovado, mas também pode ser reprovado em conjunto com outro termo, como pode ser visto nos componentes D e F que representam o uso da solução E2. Quando um termo é aprovado, a borda do *card* fica da cor verde, e quando reprovado, a borda fica da cor vermelha.

**Avaliação e lições aprendidas** Os problemas destacados pelo gerente de produto no protótipo anterior foram solucionados através de um design mais simples e com a possibilidade de agrupar termos para classificá-los como termos aprovados ou reprovados. Por fim, este protótipo foi aprovado para dar continuidade ao projeto.

## 5 Discussão

Este trabalho apresenta um processo evolutivo de engenharia de requisitos utilizando protótipos de sistemas de rotulagem de dados, visando atender o problema das Secretarias da Fazenda sobre a tarefa de rotulagem de milhões de produtos. Para isso, foram realizadas três iterações com os protótipos para o sistema proposto.

A cada iteração, os requisitos do sistema foram evoluídos (Tabela 2). A evolução aconteceu a partir dos requisitos iniciais apresentados na Tabela 1. Além disso, foram analisados pontos fortes e fracos de cada protótipo e a partir disso foi feita uma reflexão entre esses pontos e os requisitos iniciais, tendo como resultado os requisitos refinados. Estes requisitos refinados foram validados pelo gerente de produto e pela equipe de desenvolvimento. A Tabela 2 apresenta o ID dos requisitos iniciais da Tabela 1, os pontos fortes que possibilitaram a evolução dos requisitos e por fim, os requisitos refinados.

Apesar dos pontos fracos não estarem relacionados na Tabela 2 com os requisitos refinados, esses foram importantes no processo evolutivo dos requisitos, pois possibilitaram a reflexão sobre o que o sistema não devia fazer. Além disso, os pontos fracos evidenciaram problemas como a avaliação individual de produtos, sobrecarga de informações no protótipo e a impossibilidade de agrupar termos.

Como exemplo, o requisito RR1 foi um refinamento do requisito RI2, pois deixa claro como devem ser apresentados os produtos e suas categorias. Este refinamento foi necessário devido à ambiguidade do requisito inicial RI2. Além disso, durante a construção dos protótipos, foi observado que a visualização de

**Tabela 2.** Requisitos Refinados do Sistema de Rotulagem a partir dos requisitos iniciais e dos pontos fortes dos protótipos.

| Req. Iniciais | Pontos Fortes                                                     | Requisitos Refinados                                                    |
|---------------|-------------------------------------------------------------------|-------------------------------------------------------------------------|
| RI2           | PF5: Design <i>clean</i>                                          | <b>RR1:</b> Visualizar títulos de produtos por NCM pré-selecionado      |
| RI3           | PF2: Status de andamento                                          | <b>RR2:</b> Visualizar status de progresso de títulos rotulados         |
| RI3           | PF1: Gráfico de barra de uma palavra representativa de um produto | <b>RR3:</b> Visualizar distribuição e quantidade de títulos de produtos |
| RI1           | PF4: Destaque de termos aprovados e reprovados por cor            | <b>RR4:</b> Identificar visualmente termos aprovados e reprovados       |
| RI1           | PF6: Inclusão de grupos de termos para exclusão ou inclusão       | <b>RR5:</b> Rotular títulos por termos ou grupos de termos              |
| RI1           | PF3: Filtros de termos                                            | <b>RR6:</b> Possibilitar filtros por termos                             |

produtos por NCM proporcionava uma interface mais simples e limpa, conforme o ponto forte PF5 apontado pelo gerente de produto.

Os requisitos RR1 ao RR4 referem-se a requisitos de visualização. Esses requisitos de visualização foram adaptados de outros sistemas de rotulagem encontrados na literatura. Nesse sentido, os elementos encontrados contribuem para a melhor identificação das informações e contextualizar o usuário do que está acontecendo no sistema.

O requisito RR5 representa uma abordagem de rotular por termos ou grupos de termos. A importância desse requisito é gerada pela necessidade do usuário em avaliar muitos dados. A partir disso, foi identificado que existem muitos produtos com termos semelhantes e como solução foram criados componentes de interface que apresentassem ao usuário os termos semelhantes mais usados nas descrições dos produtos. Assim, o especialista, ao invés de classificar milhões de produtos individualmente, identifica apenas os termos mais representativos, como também os termos menos representativos que na maioria das vezes são produtos que foram classificados com o código NCM incorreto pelos contribuintes.

O requisito RR6 diz respeito a possibilidade de identificar termos através de filtros de pesquisa. Assim, os usuários podem pesquisar um termo específico, levando em consideração que esses usuários são especialistas e podem ter uma ideia prévia de quais termos devem ser analisados primeiro.

A última evolução do protótipo atendeu a todos os requisitos refinados, o que permitiu à equipe de desenvolvimento prosseguir com a construção do sistema tomando como base esses requisitos refinados.

## 6 Conclusões

Os sistemas de rotulagem estão cada vez mais sendo utilizados na indústria [3] e na academia [11,22]. A elicitação de requisitos deste tipo de sistema deve identificar os requisitos que o sistema deve ter, visando maximizar a análise humana sobre uma grande quantidade de dados.

Este trabalho descreve uma experiência com a elicitaco de requisitos de sistemas de rotulagem. O objetivo do sistema   ajudar o usu rio na inspeco de categorias de milhes de produtos em uma s viso. A partir de requisitos iniciais, foi poss vel realizar um processo de refinamento dos requisitos com tr s iteraes com o gerente de produto e a equipe de desenvolvimento.

A contribuico desta pesquisa   o processo de refinamento de requisitos durante as evolues de prot tipos, no contexto de sistemas de rotulagem. A reflexo deste ponto   referente a natureza do sistema de rotulagem, pois, apesar da necessidade, as equipes no esto acostumadas com esse tipo de desenvolvimento de sistema e isso reflete na efici ncia da etapa de identificao de requisitos.

A experi ncia apresentada est limitada a um sistema de rotulagem para o tipo de dado textual, o que implica que na pesquisa explorat ria, no foram realizadas an lises de requisitos de interface em sistemas de rotulagem para outros tipos de dados, como, por exemplo,  udio ou imagens.

Como perspectiva futura, outros sistemas de rotulagem podem ser analisados junto a outras equipes de desenvolvimento, envolvendo desde o desenvolvedor at  o gerente de projetos, visando o refinamento dos requisitos durante a evoluo da interface. Al m disso, outros tipos de requisitos podem ser investigados, como desempenho, confiabilidade, entre outros. Com esse *Know-how*, podem ser definidos procedimentos espec ficos para elicitaco de requisitos em sistemas ML.

## 7 Agradecimentos

Agradecemos   FAPEAM pelo apoio financeiro por meio do POSGRAD e dos processos de n mero 062.00638/2019 e 062.00150/2020, CNPq por meio do processo de n mero 314174/ 2020-6, e da CAPES - C digo Financeiro 001. Esta pesquisa, de acordo com o artigo 48 do Decreto n  6.008/ 2006, foi parcialmente financiada pela Samsung Electronics da Amaz nia Ltda, nos termos da Lei Federal n  8.387/1991, por meio do conv nio n  003/2019, firmado com o ICOMP/UFAM e mediante o contrato 001/2020, firmado com a UFAM e a FA-EPI.

## Refer ncias

1. Acuna, D., Ling, H., Kar, A., Fidler, S.: Efficient interactive annotation of segmentation datasets with polygon-rnn++. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 859–868 (2018)
2. Alpaydin, E.: Introduction to Machine Learning. MIT Press (2020)
3. Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T.: Power to the people: The role of humans in interactive machine learning. *Ai Magazine* **35**(4), 105–120 (2014)
4. Cui, W.: Visual analytics: A comprehensive overview. *IEEE Access* **7**, 81555–81573 (2019)
5. Dudley, J.J., Kristensson, P.O.: A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems (TiiS)* **8**(2), 1–37 (2018)

6. Felix, C., Dasgupta, A., Bertini, E.: The exploratory labeling assistant: Mixed-initiative label curation with large document collections. *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* pp. 153–164 (2018)
7. Flach, P.: *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press, Cambridge, United Kingdom (2012)
8. Gonzales, G.R.a., Horita, F.: Supporting visual analytics in decision support system: A systematic mapping study. Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3424953.3426483>
9. Governo do Estado de São Paulo: O que é e como é investido o ICMS. <https://portal.fazenda.sp.gov.br/Noticias/Paginas/0-que-%C3%A9-e-como-%C3%A9-investido-o-ICMS.aspx>, Acesso em 23 de fevereiro de 2021.
10. Huang, S.W., Tu, P.F., Fu, W.T., Amanzadeh, M.: Leveraging the crowd to improve feature-sentiment analysis of user reviews. In: *Proceedings of the 2013 international conference on Intelligent user interfaces*. pp. 3–14 (2013)
11. Kim, B., Glassman, E., Johnson, B., Shah, J.: *ibcm: Interactive bayesian case model empowering humans via intuitive interaction* (2015)
12. Kim, B., Patel, K., Rostamizadeh, A., Shah, J.: Scalable and interpretable data representation for high-dimensional complex data (2015)
13. Mannio, M., Nikula, U.: Requirements elicitation using a combination of prototypes and scenarios. In: *WER*. pp. 283–296. Citeseer (2001)
14. Mitchell, T.M.: Machine learning and data mining. *Communications of the ACM* **42**(11), 30–36 (1999)
15. Nadj, M., Knaeble, M., Li, M.X., Maedche, A.: Power to the oracle? design principles for interactive labeling systems in machine learning. *KI-Künstliche Intelligenz* pp. 1–12 (2020)
16. Nalitsnik, M., Gutman, D.A., Kong, J., Cooper, L.A.: An interactive learning framework for scalable classification of pathology images. In: *2015 IEEE International Conference on Big Data (Big Data)*. pp. 928–935. IEEE (2015)
17. Plummer, B., Kiapour, H., Zheng, S., Piramuthu, R.: Give me a hint! navigating image databases using human-in-the-loop feedback. In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. pp. 2048–2057. IEEE (2019)
18. Porter, R., Theiler, J., Hush, D.: Interactive machine learning in data exploitation. *Computing in Science & Engineering* **15**(5), 12–20 (2013)
19. Receita Federal - Ministério da Economia: NCM - Nomenclatura Comum do Mercosul. <https://receita.economia.gov.br/orientacao/aduaneira/classificacao-fiscal-de-mercadorias/ncm>, Acesso em 23 de fevereiro de 2021.
20. Roh, Y., Heo, G., Whang, S.E.: A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering* (2019)
21. Secretaria da Fazenda: Portal da Nota Fiscal Eletrônica - Informações sobre Estatísticas. <https://www.nfe.fazenda.gov.br/portal/infoEstatisticas.aspx>, Acesso em 22 de fevereiro de 2021.
22. Wallace, B.C., Small, K., Brodley, C.E., Lau, J., Trikalinos, T.A.: Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In: *Proceedings of the 2nd ACM SIGHIT international health informatics symposium*. pp. 819–824 (2012)
23. Yuan, J., Chen, C., Yang, W., Liu, M., Xia, J., Liu, S.: A survey of visual analytics techniques for machine learning. *Computational Visual Media* pp. 1–34 (2020)