Identificação de Comentários Agressivos em Vídeos Online Utilizando Técnicas Recomendação.

Marlon Miranda da Silva[1], **Ramon Figueiredo Pessoa**[2],

[1]Instituto de Informática – Pontifícia Universidade Católica
de Minas Gerais (PUC-MINAS) – MG – Brazil

[2]Instituto de Ciências Exatas e Informática

m.marlon.ms@hotmail.com, ramon.fgrd@gmail.com

Resumo. *Com o aumento da popularidade das redes sociais, problemas relacionados ao cyberbullying são cada vez mais constantes. O cyberbullying está associado a diversos atos praticados na internet com o intuito de intimidar, ofender, criticar, entre outros. Estudos indicam que pessoas que sofrem esse tipo de violência estão mais propensas a ter depressão, timidez e em casos extremos pode levar até ao suicido. Este trabalho tem como objetivo aplicar técnicas de sistemas de recomendação para o ranqueamento de comentários agressivos no YouTube. Para isso, nossa metodologia apresenta um framework com vários algorítimos de recomendação. O modelo permite que qualquer usuário que forneça feedbacks preliminares possa receber recomendações de comentários possivelmente ofensivos. O framework desenvolvido permitiu estudar qual abordagem de recomendação é mais eficaz para o domínio do problema. No final deste trabalho são descritos os resultados obtidos, onde os melhores algoritmos foram uma abordagem de sistema de recomendação baseado em usuário, atingindo cerca de 61% de precisão e a técnica de aprendizado de máquina Support Vector Machine que atingiu cerca de 61% de acurácia.*

1. Introdução

Com o aumento desenfreado do uso da internet e redes sociais observa-se o crescimento de diversos resultados negativos, entre eles o cyberbullying [Tanaka 2015]. O bullying, que pode ser visto como "um comportamento consciente, intencional, deliberado, hostil e sistemático, de uma ou mais pessoas, cuja intenção é ferir os outros" [Souza et al. 2014]. De acordo com Martins [Martins 2005], o bullying pode se manifestar de três formas. Na primeira forma, assume-se os comportamentos físicos que podem se desdobrar em várias formas de agressões. A segunda forma de bullying envolve comportamentos de cunho verbal, como insultos e comentários racistas, homofóbicos ou que demostrem qualquer diferença com outra pessoa. Por último, existem os comportamentos indiretos que baseiam-se na exclusão sistemáticas de outras pessoas para favorecimento próprio. Quando esta violência é estendia para o âmbito digital denomina-se cyberbullying (ou bullying digital), em resumo, podemos definir cyberbullying como uma forma de provocação/intimidação por meio de mensagens, fotos ou vídeos em redes sociais ou páginas web comum [SANTOMAURO 2010].

A ideia do trabalho é aplicar técnicas típicas de sistemas de recomendação para à detecção de comentários agressivos em vídeos do YouTube, servindo como uma ferra-

menta que recomende comentários agressivos, auxiliando na gestão dos comentários em um vídeo online, permitindo melhor análise por parte dos responsáveis do canal, uma vez que os comentário mais agressivos serão apresentados no topo de um *ranking*. Nas seções seguintes será definido o problema e suas subdivisões evidenciando os objetivos e a abordagem que será adotada.

1.1. Motivação

Considerando que o espaço virtual possa ser uma forma de praticar atos ilícitos e o aumento expansivo de pessoas utilizando as redes sociais , atualmente é possível observar o aumento de crimes virtuais. De acordo com a pesquisa realizada pela Safernet realizada em 2009 [OLIVEIRA 2015], com 2525 alunos e 966 educadores das redes pública e privada dos estados do Rio de Janeiro, Paraíba, Pará e São Paulo, cerca de 36% dos alunos tem um amigo que já foi vítima de cyberbullying ao menos uma vez, enquanto que 26% sabem de casos de cyberbullying em sua escola, 60% dos jovens que sofreram cyberbullying são meninos. Em um inquérito online envolvendo crianças e adolescentes de 12 a 17 anos, cerca de três quartos dos participantes foram identificados com a presença de pelo menos um incidente de cyberbullying no último ano, sendo que, desse grupo, quase 85% também eram vítimas de bullying [Juvonen and Gross 2008].

De acordo com a Figura 1, as redes sociais podem contribuir com casos de racismo, como o caso da jornalista Maria Júlia Coutinho ocorrido em julho de 2015 , onde diversas pessoas fizeram citações racistas à jornalista na página do Jornal Nacional no Facebook [Paulo 2015]. Outro caso recente foi o da atriz Taís Araújo (Figura 2) que recebeu diversos comentários agressivos após postar uma foto em uma Rede Social [Rio]. Os ataques em redes sociais podem trazer diversos problemas para a saúde física e mental de crianças, jovens e adultos, entre eles estão à depressão, ansiedade, medo e até a morte em casos mais extremos [Juvonen and Gross 2008, Maidel 2009].

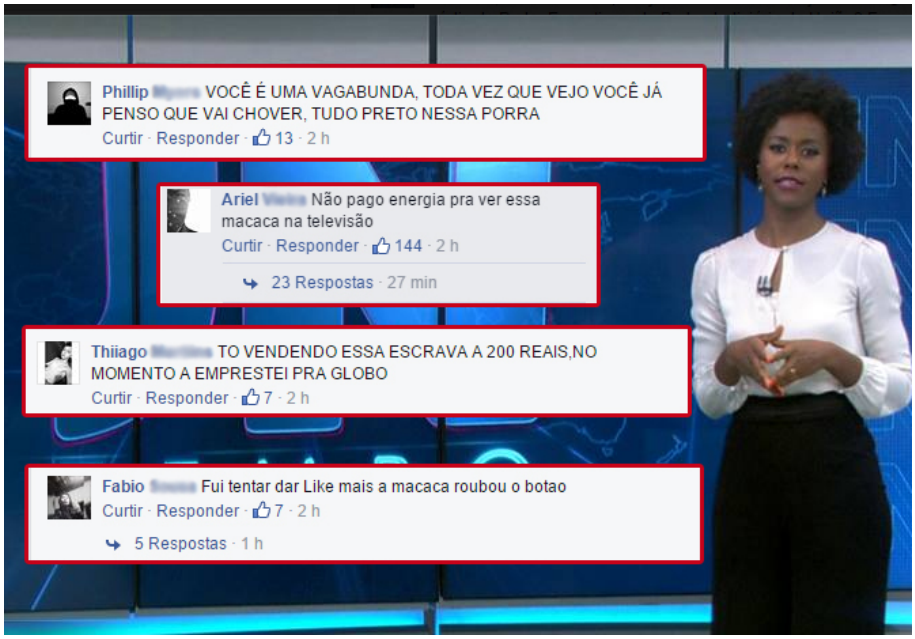

Figura 1. Caso Maju – Um crime de cyberbullying que aconteceu em julho de 2015, quando a apresentadora Maria Júlia Coutinho foi alvo de ataques racistas em redes sociais.

Figura 2. A atriz Taís Araújo é alvo de comentários racistas no Facebook em novembro de 2015.

1.2. Justificativa

Conforme citado por Brown [Brown et al. 2006], o cyberbullying mais comum ocorre por meio de comentários, que podem ser anônimos e atingir grandes quantidades em pouco tempo. Atualmente existem leis que garantem o direito de um cidadão para esse tipo de crime, como Lei Carolina Dieckman (12.737/2012) sancionada em dezembro de 2012 que prevê alterações no código penal para caso de crimes de informática. Outra lei importante para o combate do cyberbullying é a lei nº 13.185, de seis de novembro de 2015, que caracteriza como crime o cyberbullying [Pedroso and Gonçalves 2016]. Monitorar esses casos na intenção de prevalecer à lei pode ser a solução para minimizar esse tipo de violência que vem crescendo ano após ano.

1.3. Abordagem

O framework proposto neste trabalho utiliza as técnicas de sistemas de recomendação baseada em usuários, baseada em itens, não-personalizado e baseada em conteúdo para realizar o ranqueamento dos comentários. Os ranqueamentos gerados pelas técnicas baseadas em usuário e item utilizam as avaliações de outros usuários para predizer o nível de agressividade dos comentários. Já o ranqueamento não-personalizado utiliza a média das avaliações para definir o grau de agressividade dos comentários. Finalmente, o modelo baseado em conteúdo utiliza os termos mais agressivos para predizer o grau de agressividade de determinado comentário.

O trabalho proposto apresenta duas partes distintas, uma para avaliação dos comentários de determinado vídeo e outra para a avaliação e recomendação dos comentários

com base nas informações obtidas anteriormente.

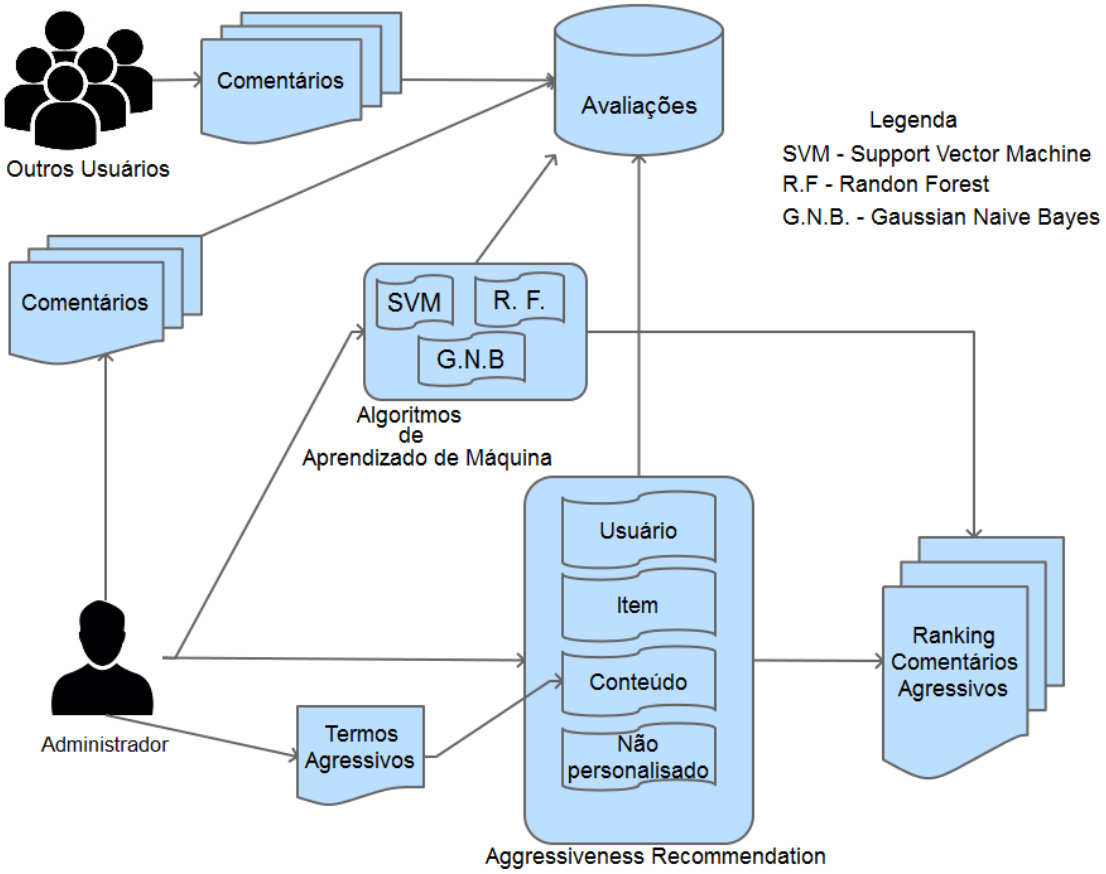

Figura 3. Fluxo de ações do Modelo Proposto

A primeira parte destina-se a possibilitar que os usuários possam avaliar os comentários e definir, com base na sua opinião, qual o nível de agressão de uma parcela da base total de comentários. Assim, os *feedbacks* obtidos são armazenados em uma base de dados.

Na segunda etapa o administrador escolhe um modelo de sistema de recomendação ou técnicas de aprendizado de máquina para obter o *ranking* de comentários agressivos. Para os modelos de recomendação ou técnicas de aprendizado de máquina, deve existir *feedbacks* pré-cadastrados pelo usuário final. Conforme mostrado pela Figura 4, sistemas de recomendação baseados em usuário utilizam as avaliações feitas pelos usuários finais para definir qual a similidade entre eles e os outros usuários, a partir dos usuários mais similares, define qual o item que mais se adequará a um usuário final, enquanto nos modelo baseado em itens os itens são recomendados aos usuários de acordo com similidade entre outros itens.

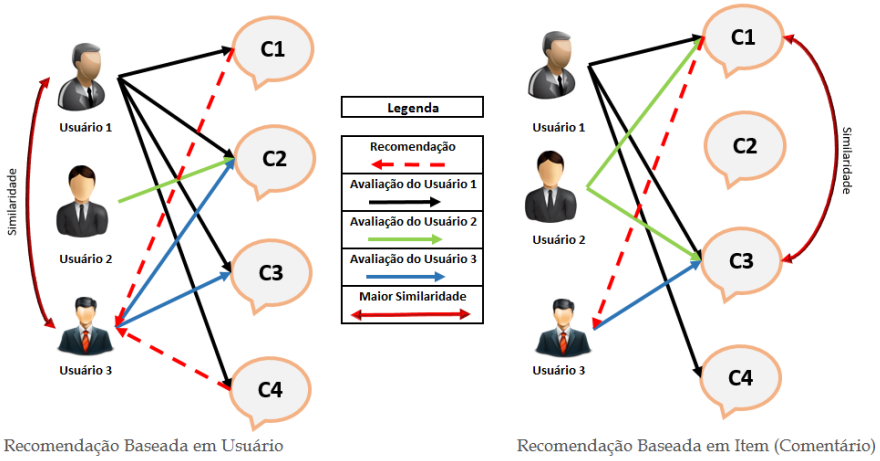

Figura 4. Exemplo de Sistema de Recomendação Baseado em Usuário e Item

O sistema de recomendação baseado em conteúdo, utiliza o vocabulário dos documentos para obter o TF-IDF (abreviação do inglês *Term Frequency–Inverse Document Frequency*, que significa frequência do termo–inverso da frequência nos documentos) e o algoritmo Rocchio. O TF-IDF, bastante utilizado em sistemas de recomendação, mede a importância de uma palavra dentro de uma coleção de documentos, no caso deste trabalho, são os comentários, levando em consideração a frequência dentro de cada comentário. A Figura 5 mostra um exemplo do algoritmo de Rocchio, este algoritmo utiliza vetores TF-IDF para representar os documentos e constrói um vetor para cada categoria juntando os vetores de documentos [Rodrigues 2016]. Desta forma os comentários são ranqueados com base na distância do cosseno, descrita na seção 2.6.2.

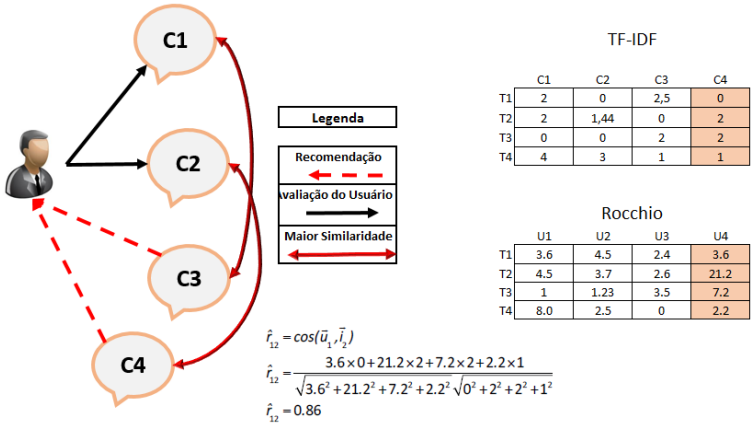

Figura 5. Exemplo do algoritmo Rocchio utilizado no Sistema Recomendação Baseado em Conteúdo

O sistema de recomendação não-personalizado utiliza somente as avaliações feitas pelos usuários para predizer determinado documento para um novo usuário. Para isso a predição é feita com base na média das avaliações feitas pelos usuários em determinado documento (Figura 6).

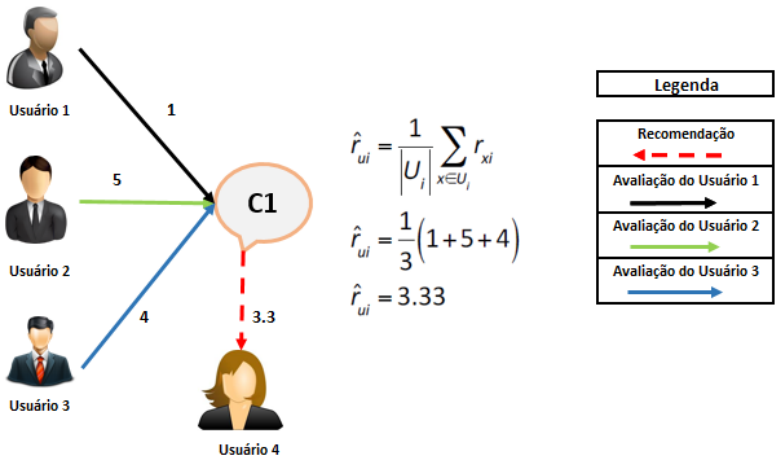

Figura 6. Exemplo de calculo realizado pelo Sistema de Recomendação Não-Personalizado

1.4. Objetivos

1.4.1. Geral

Este trabalho propõe auxiliar na detecção de agressões virtuais na rede social YouTube, tendo como objetivo apresentar um modelo de sistema de recomendação de comentários agressivos, baseado na experiência dos usuários do YouTube e definir, através de um *ranking*, os comentários que apresentam maiores níveis de agressividade. Dessa forma, as pessoas que gerenciam o conteúdo na web terão acesso à uma ferramenta que os ajude a analisar grandes volumes de dados.

1.4.2. Específicos

- Desenvolver os modelos de sistemas de recomendação baseados em usuário, item , não-personalizados e em conteúdo para fazer predições de comentários agressivos.
- Desenvolver uma plataforma de avaliação de comentários on-line.
- Utilizar técnicas de aprendizado de máquina usados na literatura para fazer a predição de comentários agressivos.
- Avaliar dentre os modelos mais eficazes para o domínio em questão do problema com base na análise de métricas comummente usados na literatura.

1.5. Organização do Texto

O trabalho esta organizado da seguinte forma: A seção 2 mostra uma revisão das principais conceitos envolvidos neste trabalho. Na seção 3 é apresentado os principais trabalhos relacionados à identificação do cyberbullying na web. A seção 4 descreve a metodologia adotada para a solução do problema em questão com as implementações dos modelos de recomendação e técnicas de aprendizado de máquina. Na seção 5 são apresentadas as métricas de avaliação dos modelos. A seção 6 apresenta aconfiguração dos experimentos utilizados. Na seção 7 é realizado a análise sobre as métricas obtidas através dos experimentos realizados. Na seção 8 são mostradas as conclusões. Finalmente na seção 9 é apresentado as direções futuras.

2. Referêncial Teórico

2.1. Cyberbullying

O termo bullying, derivado da palavra bully (do inglês "Valentão") é um termo que é utilizado para determinar o ato de "brutalizar", "tiranizar" e de modo geral maltratar de forma abusiva e constante de tal forma que a vítima se sinta amedrontada [Fante 2005]. O bullying é observado em diversos ambientes, sendo que o mais comum visto em ambientes escolares. O cyberbullying é relativamente novo da literatura e está relacionado ao uso da tecnologia para praticar o bullying [Maidel 2009]. Pesquisas para detecção de cyberbullying estão evoluindo nos últimos dez anos, porém avanços significativos são vistos nos últimos três anos com melhorias na análise de sentimentos.

O cyberbullying tem chamado bastante atenção nos últimos anos devido a gravidade em que os casos podem alcançar [Tanaka 2015]. Casos ocorridos entre os adolescentes podem ser ainda mais graves. Em agosto de 2013, Hanna Smith se suicidou após diversas ofensas na rede social Ask.fm. O culpado não foi identificado, já que o site permitia o anonimato dos usuários. De acordo com Tanaka:

> " O cyberbullying é muito mais grave que o bullying presencial, uma vez que pode gerar efeitos nocivos em lugares muito distantes, podendo atingir pessoas que o ofensor nem conheça, já que na internet, é possível enviar mensagens desonrosas a qualquer pessoa"[Tanaka 2015].

2.2. Cyberbullying em redes Sociais

À medida que a popularização e utilização de redes sociais aumentam os casos de cyberbullying são cada vez mais evidentes. Nos EUA mais da metade dos adolescentes já foram vítimas de pelo menos um tipo de cyberbullying [Wikipedia 2016]. Existem várias formas de se praticar o cyberbullying, sendo que as mais comuns são através de comentários que pode chegar a atingir um público infinito e em pouco tempo [Brown et al. 2006]

A violência praticada em redes sociais como o Facebook pode causar danos severos à subjetividade dos usuários, como medo, vergonha e depressão. O caso de Megan Meier, uma adolescente estadunidense, ficou mundialmente famoso. Megan, aos treze anos, cometeu suicídio depois de sofrer cyberbullying [Maag 2007].

2.3. Cyberbullying no Brasil

No contexto brasileiro uma investigação realizada em 2010 apontou que aproximadamente 30% dos estudantes do ensino fundamento foi vítima de bullying [Malta et al. 2010]. O conceito de cyberbullying não se restringe somente ao ambiente escolar, várias celebridades como as atrizes Carolina Dkermarn e Thais Araújo já sofreram de cyberbullying [Rio , KOHN 2012].

2.4. Extração de Dados Em Redes Sociais

Conforme citado por Santos [SANTOMAURO 2010], a web é atualmente o maior repositório de informações existentes no mundo. Com a web é possível compartilhar grandes massas de dados, que nem sempre são utilizadas em seu potencial máximo. Neste

contexto Rezende [Rezende et al. 2003] afirma que: “Devido à incapacidade do ser humano de interpretar tamanha quantidade de dados, muita informação e conhecimento, possivelmente úteis, podem estar sendo desperdiçados, ficando ocultos dentro das bases de dados espalhadas pelo mundo” [Rezende et al. 2003].

Levando em consideração a grande capacidade de informação que a web pode produzir e a grande presença dos usuários em redes sociais como Facebook e Twitter, a mineração de dados pode ser uma grande aliada no estudo do comportamento e predição de padrões sociais [Poloni and Tomaél 2014].

2.5. Recuperação da Informação

A informação possui cada vez mais importância para a sociedade de forma geral ano após ano [Chaves 2014]. A quantidade de informações existente no mundo vem desafiando tanto aqueles que precisam encontrá-las quanto os encarregados de organizá-las, para que seja possível uma recuperação bem-sucedida [Maia and Souza 2008].

A recuperação da informação tem como um de seus desafios atender às necessidades específicas do usuário de forma rápida e precisa. A maioria dos métodos atualmente utilizados na recuperação da informação usa termos formados por palavras chaves como unidade básica de acesso à informação [Souza et al. 2014]. Atualmente a maior fonte de dados que possuímos é a web, porém, a web é constituída de características não padronizadas [Souza et al. 2014].

2.6. Sistemas de Recomendação

Diversas aplicações buscam orientar os seus usuários a navegar de forma efetiva através dos itens de possível interesse, por exemplo, empresas como a Amazon, Netflix e Spotfy utilizam destas técnicas para recomendar conteúdo relevante para os usuários, entretanto estas aplicações podem recuperar conteúdo irrelevante para o usuário [Adomavicius and Tuzhilin 2005]. Os sistemas de recomendação buscam minimizar este tipo de problema utilizando das características do usuário para fornecer conteúdo relevante [Cazella et al. 2009]. Neste contexto, os sistemas de recomendação tentam prever os conteúdos adequados para o usuário, evitando a sobrecarga de informação [Garin et al. 2006]. Entre os modelos de sistemas de recomendação existentes destacamos os modelos baseados em usuário, item, conteúdo e não personalizado.

2.6.1. Sistemas de recomendação baseados em filtragem colaborativa

O sistemas de recomendação baseados em filtragem colaborativa dividem-se em duas abordagens distintas, baseado em usuário e baseado em item. A recomendação baseada em usuário baseia-se na recomendação dos itens com base na similidade dos itens avaliados entre dois ou mais usuários. Para tal, os usuários devem avaliar os itens do sistema. Essas avaliações permitem que os sistemas de recomendação identifiquem padrões e, por sua vez, recomendem itens de forma automática. Já a abordagem baseado no item é o inverso da baseado em usuário, ela utiliza a similaridade entre as avaliações dos itens em relação aos usuários, recomendando o item para o usuário.

Em ambas abordagens a técnica consiste em (i) calcular a similaridade do usuário/item alvo em relação aos outros usuários/itens; (ii) selecionar o grupo de

usuários/itens com maior similaridade; (iii) realizar a predição dos itens/usuários com base na na fórmula *Mean-Centering* (descrita abaixo) , onde $sim(\vec{u}, \vec{v})$representa a similaridade entre dois usuários/itens e r_{vi} a avaliação dada para o item/usuário. [Cazella et al. 2010, Costa et al. 2013].

$$\bar{r}_{u\,i} = \frac{\sum_{v=N} sim(\vec{u}, \vec{v}) \times r_{vi}}{\sum_{v=N} |sim(\vec{u}, \vec{v})|}$$

2.6.2. Sistemas de Recomendação baseados em Conteúdo

Sistemas de recomendação baseados em conteúdo recomenda aos usuários itens que possuam conteúdo semelhante aos itens que ele avaliou anteriormente. A recomendação pode ser feita através do conteúdo de *tags* que descrevem o item ou no caso deste trabalho no próprio item, uma vez que o mesmo é composto por texto. [Cazella et al. 2010, Costa et al. 2013]. Neste trabalho utilizou o TF-IDF da lista de palavras consideradas agressivas para definir o conteúdo dos comentários e a matriz Rocchio que fornece pesos aos *feedbacks* (votos) dos usuários com base nestes termos. Para calcular a matriz Rocchio foi utilizado a formula abaixo, onde R_u representa os itens avaliados por todos os usuários u e r_{uj} representa avaliação do j-ésimo usuário u para o item J.

$$\vec{U} = \frac{1}{|R_u|} \sum r_{uj} \vec{J}$$

Para calcular o valor da predição foi utilizado a distância do cosseno. A distância do cosseno ($C_{a\,b}$) calcula o ângulo formado por dois vetores que representam as avaliações dos itens na abordagem baseado no usuário ou dos usuários na abordagem baseado no item. O valor do cosseno calculado, que varia de zero a um, indica a similaridade entre os vetores. Quanto mais próximo a um, mais similares são os vetores, logo, quanto mais próximo a zero, menos similares [Costa et al. 2013]. Para normalizar a predição o valor obtido é normalizado multiplicado por 5, valor máximo da avaliação feita pelos usuários.

$$C_{a\,b} = \frac{\sum_{i=1}^{m} (W_{a\,i} \times W_{b\,i})}{\sum_{v=N}^{m} (W_{a\,i})^2 \times \sum_{i=1}^{m} (W_{b\,i})^2}$$

2.6.3. Sistemas de Recomendação Não Personalizado

Esta abordagem utiliza a média das avaliações de um item para recomendar determinado item para determinado usuário. O cálculo da predição é dado pelo valor $\bar{r}$, onde $|U_i|$ representa o número total de predições para determinado item e r_{xi} o valor avaliado para o item i por determinado usuário x. Esta abordagem apresenta alguns problemas como o fato das predições serem as mesmas para todos os usuários, outro ponto negativo é o fato das predições ignorarem o conteúdo dos comentários.

$$\bar{r} = \frac{1}{|U_i|} \sum r_{xi}$$

2.7. Técnicas de Aprendizado de máquina

Existem diversos algoritmos de aprendizado de máquina na literatura, este trabalho aborda os algoritmos *Support Vector Machine (SVM)*, *Random Forest* e *Gaussian Naive Bayes*, nos quais são descritos nas seções subsequentes.

2.7.1. Support Vector Machine (SVM)

Support Vector Machine é uma técnica de *machine learning* para classificação de textos. Nesta técnica os documentos são representados como pontos num espaço vetorial, onde as dimensões definidas por vetores de características criados a partir dos dados. A base desta técnica é encontrar um hiperplano ótimo com base no treino dos vetores de documentos, como mostrado na Figura 7 [Rodrigues 2016].

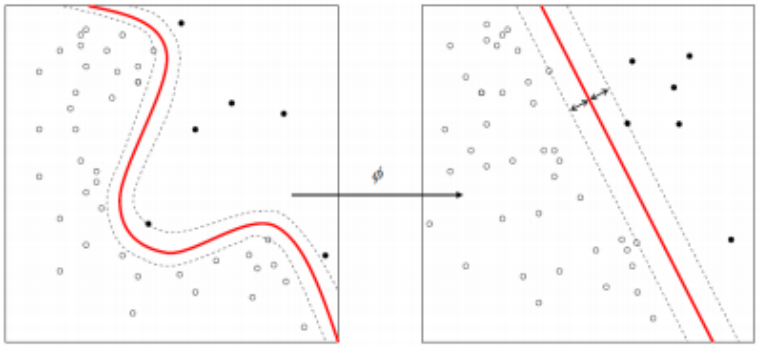

Figura 7. Hiperplano que separa 2 classes [Rodrigues 2016].

2.7.2. Random Forest

A técnica *Radom Forest* combina árvores de decisão individuais em conjunto. O algoritmo atua criando árvores binárias, nas quais, cada nó gera apenas dois nós filhos. Cada árvore vota numa determinada classe a classe mais votada é escolhida como saída do classificador. Cada árvore recebe o mesmo peso nas votações, o que significa que não há arvores que possuam valor mais importante que o valor das demais [Rodrigues 2016].

2.7.3. Gaussian Naive Bayes

Gaussian Naive Bayes é uma técnica de aprendizagem supervisionada probabilística baseada no Teorema de Bayes. Nesta técnica, assumi-se que a existência de um tipo de características não depende da existência de outras características, logo, necessita de uma base de treino já rotulada para construir um modelo de aprendizagem. Os novos documentos serão classificados com base na probabilidade de pertencerem ou não à uma classe pré-definida [Rodrigues 2016].

3. Trabalhos Relacionados

Encontramos na literatura diversos trabalhos relacionados com o cyberbullying em redes sociais. Os trabalhos utilizam como principal fonte de dados as API's disponibilizada pelas redes sociais. O termo "API" é um acrônimo de *"Application Programming Interface"*, e são associadas a biblioteca de software. As API's descrevem o funcionamento da biblioteca, em geral as API's disponibilizadas por redes sociais possibilitam a extração de grande quantidade de informação públicas de forma rápida e ágil.

Existem problemas ao utilizar as API's das redes sociais como sistemas de recuperação de informação, uma delas é a pouca precisão, trazendo muitas informações que não dizem respeito ao tema inicial da busca, ou até mesmo não possibilitando fazer um pré-filtro para a obtenção dos dados. A busca se torna ainda mais complexa quando envolve análise de sentimento, devido ao fato de conter características implícitas ou explicitas conforme citado por Ferreira [Ferreira 2010]. Para amenizar este problema, Amado [Amado et al. 2009] utilizou as informações coletadas do Instagram e as filtrou por *posts* que possuem algum termo de baixo calão, a partir desses elementos utilizou-se a plataforma Crowdflower[1] para classificar os posts como contendo ou não cyberbullying.

Seguindo na mesma linha, Dadvar [Dadvar et al. 2012] utilizou como base de dados os comentários do YouTube. Cerca de 681000 posts foram classificados e tratados utilizando a ferramenta WEKA[2] e algoritmos de aprendizado supervisionado . Os comentários identificados como cyberbullying possuíam termos de baixo calão e dessa forma foi mais fácil de classificar das expressões envolvendo sarcasmo e eufemismo segundo os autores.

Outro importante fator para reconhecer padrões de cyberbullying em redes sociais é a identificação de padrões na escrita. Para tal, Garcia [Garcia et al. 2016] utiliza a ferramenta WEKA para construir uma ferramenta de aprendizagem supervisionada e detectar expressões em redes sociais com base nas informações fornecidas por estudantes. Para melhorar a classificação dos comentários.

Dinakar [Dinakar et al. 2011] utilizou clusterização para classificar cerca de 50000 comentários coletados do YouTube, para isso foi utilizado duas abordagens, na primeira o classificador foi binário, definindo os comentários em agressivos ou não, na segunda abordagem foi utilizado classificadores multiclasse para rotular os comentários com uma ou mais rótulos. Os conjuntos de dados obtidos para cada *cluster* foram divididos em 50% para treinamento 30% para validação e 20% para teste. Cada conjunto de dados foi submetido a um pré-processamento de dados para remover os termos irrelevantes. Com posse dos dados organizados três métodos de aprendizado supervisionado foram utilizados: *Repeated Incremental*, J48 e *Support-Vector Machines* (SVM). Em termos de precisão, *Repeated Incremental* foi o melhor, embora os valores de Kappa foram menores em comparação com SVM. Os altos valores de Kappa do SVM sugerem maior confiabilidade para todas os rótulos. É importante citar que não foi encontrado nenhum trabalho que utilize sistemas de recomendação no intuito de identificar cyberbullying ou agressões.

4. Metodologia

Para alcançar o objetivo proposto, este trabalho usa os seguintes passos: 1) coleta de dados, 2) criação de um framework de sistemas de recomendação e aprendizado de máquina, 3) geração de dados estatísticos dos resultados finais de classificações usando graus de agressividade.

[1]O Crowdflower é uma plataforma de avaliação de questionários onlines. Disponível em www.crowdflower.com

[2]http://www.cs.waikato.ac.nz/ml/weka/

4.1. Coleta de Dados

A coleta dos dados foi dividida em duas etapas: coleta e classificação dos termos agressivos e coleta e classificação dos comentários agressivos. Para isso, foi feito um sistema web, chamado de "Todos contra o Cyberbulluing"[3] (Figura 8). Este sistema foi responsável por realizar uma pesquisa que buscava saber o grau de agressividade de diversas palavras e comentários para um determinado usuário.

Figura 8. Site "Todos contra o Cyberbulluing" para coleta de *feedbacks* de usuários.

A Figura 9 mostra exemplo de palavras possivelmente agressivas e seus cinco níveis de agressividade que os usuário foram solicitados a classificar. Estas palavras potencialmente ofensivas foram encontradas na web brasileira usando um *crawler*.

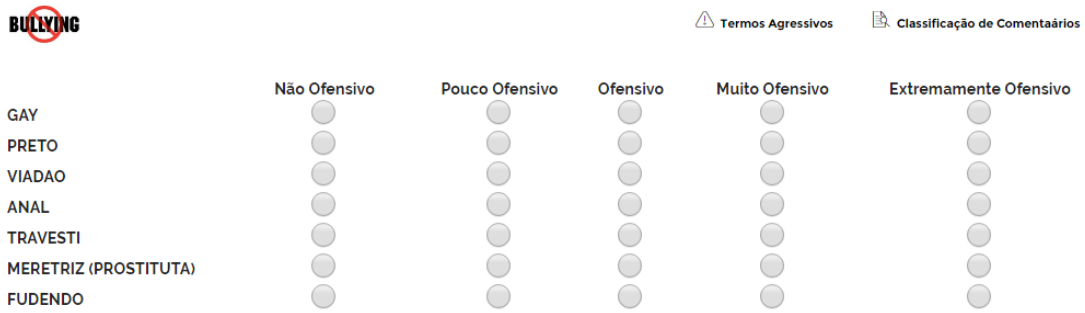

Figura 9. Formulário para a classificação de termos agressivos

A segunda etapa da coleta de dados consiste na obtenção de uma grande massa de comentários em determinado vídeo do YouTube. O vídeo escolhido foi "Felipe Neto x Marco Feliciano"[4] conforme Figura 10, este vídeo foi escolhido pois trata de diversos assuntos como política, religião e homossexualismo, temas que são discutidos no cenário atual.

[3]http://todoscontraocyberbullying.esy.es/
[4]https://www.youtube.com/watch?v=td4s51ghmXE

Figura 10. O vídeo do YouTube escolhido e disponível na base de dados foi "Felipe Neto X Marco Feliciano".

Um *crawler* utilizando a API do YouTube V3[5] foi usado para coletar cerca de 60 mil comentários dentre o total dos comentários. Outro motivo da escolha do vídeo foi o fato de não haver muitos vídeos com crimes de *cyberbullying* disponíveis por muito tempo em redes sociais. Até o momento, o vídeo possui mais de 120 mil comentários, entre eles elogios e críticas. Dos comentários coletados, foram selecionados aleatoriamente 1025 comentários, destes removendo elementos HTML, caracteres especiais e acentos e aplicando também os seguintes filtros para a escolha dos comentários : limite máximo de quinhentos[6] caracteres por comentário, possuir pelo menos um termo classificado como muito ofensivo ou extremamente ofensivo na etapa de classificação das palavras. O *feedback* dos usuários para os comentários foi feito em uma escala de 1 a 5, onde 1 é não agressivo e 5 é extremamente agressivo. Exemplos de comentários e seus cinco níveis de agressividade são mostrados na Figura 11.

[5]https://developers.google.com/youtube/v3/

[6]Este filtro foi aplicado para não ficar muito exaustivo a classificação por parte dos usuários

Figura 11. Exemplos de comentários e seus cinco níveis de agressividade.

4.2. Implementação do framework

Para a implementação do modelo de sistema de recomendação optou-se pela linguagem Java, devido a facilidade de desenvolvimento do modelo e por não encontrar na a literatura nenhuma implementação que utiliza-se esta linguagem. O sistema foi desenvolvido orientado pelo padrão *Model–view–controller* (*MVC*), sendo divido em três módulos principais: Calculadora, responsável por realizar todos os cálculos do sistema, Domínio, entidades do sistema e controladora e Útil, módulo responsável pelos métodos auxiliares como leitura e escrita de arquivos. O código está disponível em https://github.com/MarlonMiranda/AgressiveRecomendation.git.

O sistema utiliza como entrada quatro arquivos no formato .csv separado por ";". O primeiro arquivo é o *itens.csv*, contendo o código do item, neste trabalho tratado como comentário e as demais informações dos itens, o segundo arquivo é *users.csv* contendo o código do usuário e as demais informações do usuário, o terceiro arquivo é o palavras.csv, contendo as palavras classificadas como agressivas e por fim o *ratings.csv*, contendo o código do usuário, o código do item e a avaliação dada para o item. Como saída o sistema pode gerar vários tipos de medições, sendo os principais a precisão e revocação do ranqueamento criado, métricas de erro e métricas de precisão

4.3. Ranqueamento dos comentários com base no grau de agressividade

Para realizar o ranqueamento foi implementado quatro algoritmos de recomendação descritos na seção 2.6, este estudo optou-se pelo sistema de recomendação baseado em usuário, sistema de recomendação baseado no item, sistema de recomendação baseado no conteúdo e sistema de recomendação não- personalizado. Também foi usado três técnicas de aprendizado de máquina, são elas *Support Vector Machine (VM)*, *Random Forest*, *Gaussian Naive Bayes*.

5. Métricas

Todas as predições foram feitas usando os 50 comentários classificados por usuários, onde 25 comentários aleatórios foram selecionados para treino e 25 comentários aleatórios foram usados para teste. Para as técnicas de aprendizagem de máquina foram usadas técnicas de validação cruzada para evitar viés nos classificadores.

5.1. MAE

A métrica *Mean Absolute Erro* (*MAE*) representa o desvio médio entre as predições do sistema de recomendação e as avaliações feitas pelos usuários, a diferença entre elas é dado como o erro da predição , onde $|e_t| = |y_i - f_i|$, sendo y_i o i-ésimo valor recomendado e f_i o i-ésimo valor já avaliado pelo usuário [Cazella et al. 2009].

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t|$$

5.2. MSE

A métrica *Mean Squared Error* (*MSE*) apresenta a média dos quadrados dos erros, ou perda quadrática. Quanto menor o erro médio quadrático mais os valores recomendados estão próximos dos valores avaliados pelo usuário em um sistema de recomendação , onde $|e_t| = |y_i - f_i|$, sendo y_i o i-ésimo valor recomendado e f_i o i-ésimo valor já avaliado pelo usuário [Resnick and Varian 1997].

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$$

5.3. RMSE

A métrica *Root Mean Squared Error* (*RMSE*) é dada pela raiz quadrada da média divida pela média dos quadrados de todos os erros. O uso desta métrica ajuda a identificar o erro médio absoluto, penalizando erros maiores, onde $|e_t| = |y_i - f_i|$, sendo y_i o i-ésimo valor recomendado e f_i o i-ésimo valor já avaliado pelo usuário [Chai and Draxler 2014].

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$$

5.4. MAP

A métrica *Mean Average Precision* (*MAP*) mede a precisão média para um conjunto de consultas, onde cada consulta é dada pelos valores $P@n$ para todos os documentos relevantes, onde N é o número de documentos recuperados, e $rel(n)$ é uma função binária sobre a relevância do n-ésimo documento. Sendo assim, o valor do *MAP* é dado pela média dos valores *AP* de todas as consultas [Barth , Liu et al.].

$$AP = \frac{\sum_{n=1}^{N} P@n \times rel(n)}{r_q}$$

5.5. F1

A métrica *F Score* (*F1*) mede a precisão de um modelo. Esta métrica considera tanto a precisão quanto a revocação de um modelo para calcular a pontuação. Esta medida pode ser considerada como uma ponderação entre a revocação e a precisão, atingindo valores 0 e 1, onde 0 representa os piores resultados [Resnick and Varian 1997].

$$F1 = \ 2 \times \frac{precision \times recall}{precison + recall}$$

6. Configuração Experimental

Nesta seção descrevemos e planejamentos dos experimentos.

6.1. Base de Dados Utilizado

A base de dados criada neste trabalho possui 188 palavras potencialmente ofensivas, o que gerou 5194 classificações de palavras, com média de 28 classificações por palavras. Para cada usuário foi solicitado classificar 25 palavras aleatórias, totalizando aproximadamente 208 usuários participantes. Para os 1025 comentários potencialmente ofensivos, obteve-se média de 9 classificações por comentário, onde o número total de comentários classificados foi 10400. Para cada usuário foi solicitado classificar 50 comentários aleatórios, totalizando 208 usuários participantes. A base está disponível online[7]. A Tabela 1 mostra um resumo do banco de dados de classificação de palavras e comentários potencialmente ofensivas criado neste trabalho.

Categorização	**Total**
Palavras potencialmente ofensivas	188
Total de classificação de palavras	5194
Média de classificações por palavras	28
Usuários que classificaram palavras	208
Palavras classificadas por cada usuário	25
Vídeos no banco de dados	1
Comentários coletados em um vídeo	63543
Comentários após o filtro de palavras ofensivas	1025
Total de classificação de comentários	10400
Média de classificações por comentário	9
Usuários que classificaram comentários	208
Comentários classificados por cada usuário	50

Tabela 1. Resumo do banco de dados de classificação de palavras e comentários potencialmente ofensivas criado neste trabalho.

A Figura 12 mostra uma sumarização das classificações de palavras mais ofensivas feita pelos usuários. Por exemplo, a palavra "prostitua" foi classificada 85 vezes como extremamente agressiva e 72 vezes como muito agressiva. A Figura 13 mostra a quantidade de votos para comentários "Extremamente agressivos", "Muito Agressivo", "Agressivo", "Pouco Agressivo" e "Não agressivo" para todas as classificações de comentários. Por exemplo, os usuários votaram 1244 vezes na opção extremamente agressivo.

[7]https://1drv.ms/u/s!AuUWmMkGikrU0QlgLFqtd8W5wJBg

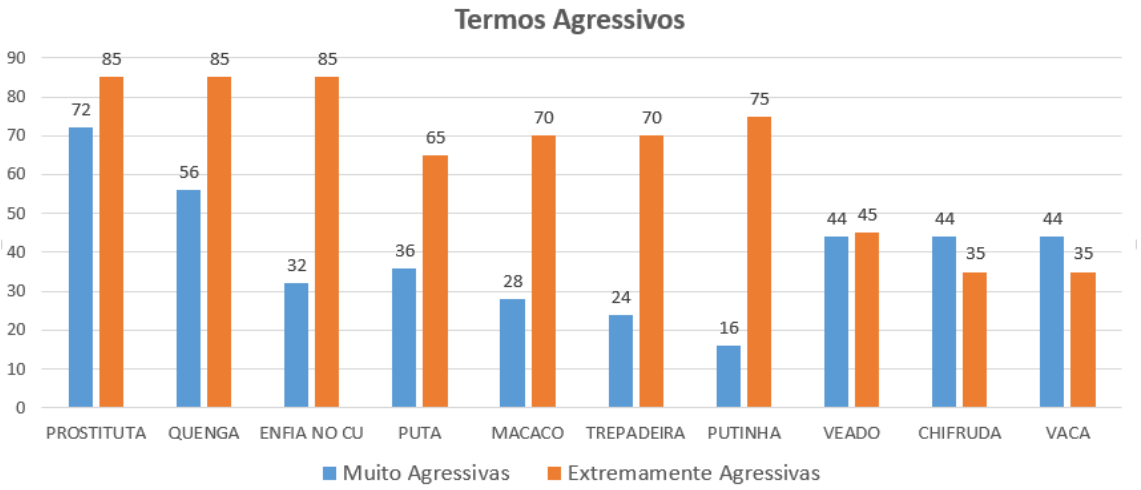

Figura 12. Sumarização das classificações de palavras mais ofensivas feita pelos usuários.

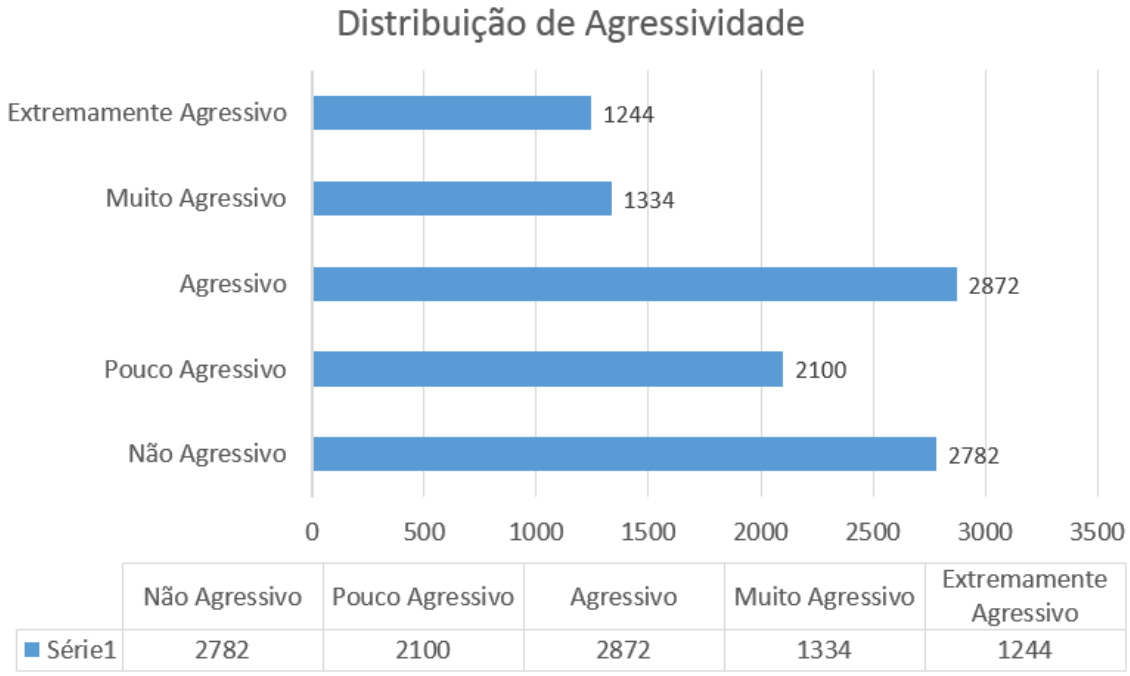

Figura 13. Quantidade de votos para comentários "Extremamente agressivos", "Muito Agressivo", "Agressivo", "Pouco Agressivo" e "Não agressivo" para todas as classificações de comentários.

6.2. Configuração utilizada nos sistemas de recomendação

Para avaliar as técnicas desenvolvidas foi aplicado um projeto fatorial 2^k [Bukh and Jain 1992], cujo o intuito é avaliar quais são os fatores que mais influenciam no resultado. Foram utilizados os seguintes fatores: (i) número de vizinhos (K-NN); (ii) diferencial mínimo relevante. O número de vizinhos implica quantos usuários/itens semelhantes serão utilizados para definir a predição. Já o diferencial mínimo relevante

refere-se ao valor aceitável pelo erro entre a predição e a avaliação do usuário, este fator define se o documento recuperado é relevante ou não.

Os experimentos foram executados variando o fator um em 2 e 30 e o fator dois entre 1 e 2. Os resultados obtidos mostram que o fator dois influencia em 99,79% do resultado final. Desta forma as métricas obtidas para sistemas de recomendação baseados em usuário e item foram executados utilizando a configuração de 30 vizinhos mais próximos e 2 para o máximo de erro considerável.

Para os sistema de recomendação baseado em conteúdo, não foi considerado todo o vocabulário da base de comentários. A utilização desta se mostrou muito demorada. Foi utilizada a base de palavras coletadas descritas na seção 4.1, totalizando 188 palavras. Este sistema utiliza somente o diferencial mínimo relevante para definir os documentos relevantes recuperados, para este fator foi utilizado o valor 2.

Para o sistema de recomendação não-personalizado, que utiliza somente o diferencial mínimo relevante para definir os documentos relevantes recuperados, foi utilizado o valor 2.

6.3. Configuração utilizada nos sistemas de aprendizado de máquina

Neste trabalho foi escolhido uma biblioteca para aprendizado de máquina em Python chamada Scikit-Learn[8] onde foi possível utilizar classificadores já conhecidos nas áreas de machine learning, sendo eles: Support Vector Machine (SVM), Random Forest e Gaussian Naive Bayes. Todos os três classificadores usaram 50% dos comentários escolhidos de forma aleatória para treino e os outros 50% foram usados para teste. A técnica de validação cruzada foi aplicada para evitar viés nos classificadores. Além disso, os algoritmos de classificação usaram os valores de parâmetros padrões definidos na biblioteca Scikit-Learn. Os parâmetros utilizados foram:

Parâmetros Support Vector Machine (SVM): $SVC(C = 1.0, cache_size = 200, class_weight = None, coef0 = 0.0, decision_function_shape = None, degree = 3, gamma =' auto', kernel =' rbf', max_iter = -1, probability = False, random_state = None, shrinking = True, tol = 0.001, verbose = False)$

Parâmetros Gaussian Naive Bayers: $GaussianNB(priors = None)$

Parâmetros Random Forest: $RandomForestClassifier(bootstrap = True, class_weight = None, criterion =' gini', max_depth = None, max_features =' auto', max_leaf_nodes = None, min_impurity_split = 1e - 07, min_samples_leaf = 1, min_samples_split = 2, min_weight_fraction_leaf = 0.0, n_estimators = 10, n_jobs = 1, oob_score = False, random_state = None, verbose = 0, warm_start = False)$

7. Resultados Experimentais

Os resultados das técnicas de sistemas de recomendação são mostrados na Tabela 2. A Tabela 2 mostra o resultado para cada métrica de erro e para a métrica *F1 Score*

[8]http://scikit-learn.org/stable/

e *Mean Average Precision (MAP)*. Por exemplo, a técnica de sistema de recomendação baseado em usuário teve a erro MAE de 0,57 numa escala de 1 a 5 e uma precisão F1 de 61.62%.

Abordagem	MAE	MSE	RMSE	MAP	F1
Usuário	0,5745	0,8550	0,7444	0,0382	61,62%
Conteúdo	1,6482	3,9295	1,9304	0,0101	39,43%
Item	1,1027	1,8674	1,2132	0,0043	54,20%
Não Personalizado	0,6723	0,8104	0,8575	0,0158	54,76%

Tabela 2. Resultados de métricas de erro e métrica F1 Score para as técnicas de sistemas de recomendação.

Os resultados foram validados estatisticamente usando uma confiança de 95% e os melhores algoritmos foram as técnicas de sistema de recomendação baseado em usuário e não-personalizado, que se mostram superiores às técnicas baseado em item e baseado em conteúdo quando comparados pela métrica F1 (precisão), entretanto não é possível afirmar que a técnica baseado em usuário é superior que a técnica baseada em conteúdo para a métrica F1, uma vez que, seu intervalo de confiança envolve o zero, portanto, não é possível afirmar que eles são estatisticamente diferentes.

Método comparado	Limite Inferior	Limite Superior
Não Personalizado	-0,0042	0,0079
Item	0,0396	0,0596
Conteúdo	0,0634	0,0867

Tabela 3. Intervalo de confiança da comparação entre sistema de recomendação baseado em usuário e as demais abordagens para a métrica F1.

A métrica de erro RMSE, que é uma métrica que penaliza erros maiores, permite uma melhor análise dos erros. Portanto, com base na Tabela 2 podemos afirmar que a técnica baseada em usuário possui melhor precisão e consequentemente menor erro. E com 95% de confiança, podemos afirmar que a técnica baseada em usuário se mostrou melhor do que as demais abordagens, uma vez que, o intervalo de confiança da comparação não inclui o zero, conforme mostrado na Tabela 4 .

Método comparado	Limite Inferior	Limite Superior
Não Personalizado	-0,2185	-0,0077
Item	-0,4873	-0,3217
Conteúdo	-1,2711	-1,1010

Tabela 4. Intervalo de confiança da comparação entre o sistema de recomendação baseado em usuário e as demais abordagens para a métrica RMSE.

A melhor técnica de aprendizado de máquina foi a *SVM* com 61,43% de acurácia. Um teste pareado estatístico (com 95% de confiança) também foi feito usando o *Support Vector Machine* comparado com o *Random Forest* ou *Gaussian Naive Bayes* e *Support Vector Machine* mostrou superior aos outros métodos. Os resultados das técnicas de aprendizagem de máquina são mostrados na Tabela 5.

Técnica	**Acurácia**
Support Vector Machine (*SVM*)	61,43 %
Random Forest	52,55%
Gaussian Naive Bayes	54,20%

Tabela 5. Resultados de acurácia para as técnicas de aprendizagem de máquina.

8. Conclusão

Este trabalho apresentou duas soluções para detecção de cyberbullying em comentários online feitos em vídeos e também construiu um banco de dados com 10400 classificações de comentários feitas por 208 usuários, disponível online conforme seção 6.1. O código da implementação das quatro técnicas de sistemas de recomendação etá disponível online, conforme seção 4.2 A validação foi feita separando as 10400 classificações em conjuntos de treino e teste e predizendo o possível nível de agressividade que um usuário iria fornecer baseado no conhecimento que o sistema construiu usando o conjunto de treino.

O trabalho atua em soluções não encontradas na literatura ao mapear o problema usando técnicas de sistemas de recomendação. A abordagem proposta é diferente das soluções encontradas na literatura que comumente usam somente técnicas de aprendizado de máquina para tentar solucionar o problema. Vários trabalhos encontrados na literatura usam principalmente o algoritmo *Support Vector Machine (SVM)*, porém este trabalho mostrou que o algoritmo de aprendizado de máquina *SVM* foi superior à técnica de *Random Forest* e *Gaussian Naive Bayes* na base de dados usada. Testes estatísticos, com 95% de confiança, mostram que o uso de sistemas de recomendação para predizer comentários potencialmente ofensivos é uma boa solução, pois propõe um sistema mais dinâmico que recomenda possíveis casos de cyberbullying para um usuário específico.

A melhor solução usando sistemas de recomendação foi o sistema de recomendação baseado em usuário. Usando a configuração 30 vizinhos mais próximos (k=30), média dos pesos de *feedbacks* feitos por usuários, normalização da média do usuário e similaridade do cosseno. Uma possível justificativa para a abordagem ser melhor do que a abordagem baseado em item pode ser explicado porque temos mais classificações de comentários feitas por usuário (total de 50) do que a quantidade de avaliações feitas para cada comentário (média de 9 avaliações por comentários). Como a técnica baseado em usuário usa a similaridade de usuários e a técnica baseado em item usa a similaridade de itens, uma maior quantidade de avaliações de comentários por parte dos usuário beneficiam a primeira técnica. Já a melhor técnica de aprendizado de máquina foi a técnica de *Support Vector Machine* . Testes estatísticos, com 95% de confiança, mostra que o algoritmo *Support Vector Machine* foi superior ao *Random Forest* e ao *Gaussian Naive Bayes*.

Com base na métrica F1 observa-se que o algoritmo que obteve melhor performance foi o sistema de recomendação baseado por usuário no qual teve aproximadamente 61% de sucesso. O sistema de recomendação baseado em item não obteve um bom desempenho pois o número de avaliações individuais de cada item estava na média de 9 avaliação inferior quando comparado com a classificação baseado no usuário, onde a mesma teve em média 49 avaliações. O sistema de recomendação baseado em conteúdo obteve a pior média, isso deve-se possivelmente à dois fatores importantes, o primeiro deles é o fato do algoritmo ter levado em consideração somente as palavras identificadas como agressivas, o segundo fator está relacionado com domínio de linguagem dos comentários, são comentários escritos por pessoas aleatórias, ou seja, de diversas idades e formações, não contendo nenhum tipo de padrão de linguagem. O sistema de recomendação não personalizado obteve a segunda maior média com aproximadamente 54% precisão, demostrando que para esse caso a média das avaliações pode ter grande influência no resultado final, sendo uma ótima alternativa para o problema de *Cold Start*[9].

Uma possível limitação da nossa abordagem é a dependência e necessidade de coleta de *feedbacks* por parte dos usuários. Este é um problema relacionado a pesquisas em sistemas de recomendação. A medida que temos mais dados coletados dos usuários, melhor é o uso de técnicas de filtragem colaborativa como as técnicas baseadas em usuário e baseada em itens.

9. Trabalhos Futuros

Como trabalhos futuros, é planejado: 1) encontrar quais as melhores configurações de parâmetros das técnicas de sistemas de recomendação usadas, assim como encontrar as melhores configurações de parâmetros usados nos algoritmos de aprendizado de máquina, 2) criar e validar um algoritmo híbrido de sistema de recomendação que combina as técnicas testadas individualmente neste trabalho, 3) testar todos os algoritmos avaliados neste trabalho em bancos de dados disponibilizados por trabalhos da literatura, 4) usar novas técnicas de aprendizado de máquina, testando principalmente uma abordagem de *deep learning* que atualmente se mostra um estado da arte, 5) criar um modelo de recomendação utilizando informações como sexo e idade para melhorar o ranqueamento dos comentários.

Referências

Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749.

Amado, J., Matos, A., Pessoa, T., and Jäger, T. (2009). Cyberbullying: um desafio à investigação e à formação. *Revista Interacções*, pages 301p–326p.

Barth, F. J. Uma breve introdução ao tema recuperação de informação.

Brown, K., Jackson, M., and Cassidy, W. (2006). Cyber-bullying: Developing policy to direct responses that are equitable and effective in addressing this special form of bullying. *Canadian Journal of Educational Administration and Policy*, 57:1–36.

[9]O Cold Start, é um problema comum em sistemas de recomendação, ocorre quando os dados (avaliações) sobre os itens ou usuários não estão disponíveis no sistema, ou são escassos.

Bukh, P. N. D. and Jain, R. (1992). The art of computer systems performance analysis, techniques for experimental design, measurement, simulation and modeling.

Cazella, S. C., Nunes, M., and Reategui, E. B. (2010). A ciência da opinião: Estado da arte em sistemas de recomendação. In *XXX Congresso da Sociedade Brasileira de Computação—Jornada de Atualização em Informática (JAI)*, pages 161–216.

Cazella, S. C., Reategui, E. B., Machado, M., and Barbosa, J. L. V. (2009). Recomendação de objetos de aprendizagem empregando filtragem colaborativa e competências. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 1.

Chai, T. and Draxler, R. (2014). Root mean square error (rmse) or mean absolute error (mae)? *Geoscientific Model Development Discussions*, 7:1525–1534.

Chaves, R. S. (2014). Análise em agrupamentos de documentos eletrônicos. *Projetos e Dissertações em Sistemas de Informação e Gestão do Conhecimento*, 3(2).

Costa, E., Aguiar, J., and Magalhães, J. (2013). Sistemas de recomendação de recursos educacionais: conceitos, técnicas e aplicações. *Jornada de Atualização em Informática na Educação*, 1(1).

Dadvar, M., de Jong, F. M., Ordelman, R., and Trieschnigg, R. (2012). Improved cyberbullying detection using gender information.

Dinakar, K., Reichart, R., and Lieberman, H. (2011). Modeling the detection of textual cyberbullying. *The Social Mobile Web*, 11:02.

Fante, C. (2005). *Fenômeno bullying: como prevenir a violência nas escolas e educar para a paz*. Verus Editora.

Ferreira, E. d. B. A. (2010). Analise de sentimento em redes sociais utilizando influência das palavras. *Trabalho de Graduação-Universidade Federal de Pernambuco-UFPE. Departamento de Ciênciada Computação*.

Garcia, A. R. M., Arriaga, S. O., and Amaral, I. O. (2016). *O Uso da Internet, o Bullying, o Cyberbullying e o Suporte Social em jovens do 3º Ciclo–Um Estudo não Experimental Correlacional realizado numa Escola Portuguesa.* PhD thesis, ISMT.

Garin, R. S., Lichtnow, D., Palazzo, L. A. M., Loh, S., Kampff, A. J. C., Primo, T. T., Oliveira, J. P. M. d., and Lima, J. V. d. (2006). O uso de técnicas de recomendação em um sistema para apoio à aprendizagem colaborativa. *Revista brasileira de informática na educação. Vol. 14, n. 3 (set./dez. 2006), p. 49-59.*

Juvonen, J. and Gross, E. F. (2008). Extending the school grounds?—bullying experiences in cyberspace. *Journal of School health*, 78(9):496–505.

KOHN, S. (2012). Caso da carolina dieckmann serviu de alerta, diz especialista em segurança.

Liu, T.-Y., Xu, J., Qin, T., Xiong, W., and Li, H. Letor: Benchmark dataset for research on learning to rank for information retrieval. In *Proceedings of SIGIR 2007 workshop on learning to rank for information retrieval*.

Maag, C. (2007). When the bullies turned faceless. *The New York Times*, 16.

Maia, L. C. G. and Souza, R. R. (2008). Medidas de similaridade em documentos eletrnicos. *IX ENACIB*.

Maidel, S. (2009). Cyberbullying: um novo risco advindo das tecnologias digitais. *Revista electrónica de investigación y docencia (reid)*, (2).

Malta, D. C., Silva, M. A. I., Mello, F. C. M. d., Monteiro, R. A., Sardinha, L. M. V., Crespo, C., Carvalho, M. G. O. d., Silva, M. M. A. d., Porto, D. L., et al. (2010). Bullying nas escolas brasileiras: resultados da pesquisa nacional de saúde do escolar (pense), 2009. *Ciência & Saúde Coletiva*, 15(suppl 2):3065–3076.

Martins, M. J. D. (2005). O problema da violência escolar: Uma clarificação e diferenciação de vários conceitos relacionados. *Revista Portuguesa de Educação*, 18(1):93–105.

OLIVEIRA, D. (2015). Crescem denúncias de crimes na web, aponta relatório da safernet.

Paulo, G. S. (2015). Ministério público ouve suspeitos de racismo contra maria júlia coutinho. disponível em http://g1.globo.com/sao-paulo/noticia/2015/12/ministerio-publico-ouve-suspeitos-de-racismo-contra-maria-julia-coutinho.html.

Pedroso, A. M. C. and Gonçalves, D. M. (2016). Considerações sobre o bullying e ciberbullying e a proposta legal de aprimoramento ao combate à violência na escola, a partir da edição da lei nº 13.185/2015. *Seminário Nacional Demandas Sociais e Políticas Públicas na Sociedade Contemporânea*.

Poloni, K. M. and Tomaél, M. I. (2014). Coleta de dados em plataformas de redes sociais: Estudo de aplicativos. *Wirkshop de Pesquisa em Ciência da Informação*.

Resnick, P. and Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3):56–58.

Rezende, S. O., Pugliesi, J. B., Melanda, E. A., and Paula, M. d. (2003). Mineração de dados. *Sistemas inteligentes: fundamentos e aplicações*, 1:307–335.

Rio, G. Atriz taís araújo é alvo de comentários racistas em rede social. disponível em http://g1.globo.com/rio-de-janeiro/noticia/2015/11/atriz-tais-araujo-e-alvo-de-comentarios-racistas-em-rede-social.html.

Rodrigues, H. J. F. (2016). Ferramenta para text mining em textos completos.

SANTOMAURO, B. (2010). Cyberbullying: a violência virtual. *NOVA ESCOLA*.

Souza, S. B., Simão, A. M. V., and Caetano, A. P. (2014). Cyberbullying: Percepções acerca do fenômeno e das estratégias de enfrentamento. *Psicologia: Reflexão e Crítica*, 27(3):582–590.

Tanaka, C. S. I. (2015). Crimes virtuais contra a honra. *Intertem@ s ISSN 1677-1281*, 26(26).

Wikipedia (2016). Assédio virtual.