



**UNIVERSIDADE ESTADUAL DE CAMPINAS**

**FACULDADE DE TECNOLOGIA**

**CARMEN PAMELA ROSALES SEDANO**

FRAMEWORK PARA IDENTIFICAÇÃO DA SEVERIDADE DE *BULLYING* BASEADO  
EM *MACHINE LEARNING* E LÓGICA *FUZZY*

A BULLYING-SEVERITY IDENTIFIER FRAMEWORK BASED ON MACHINE  
LEARNING AND FUZZY LOGIC

**LIMEIRA**

**2017**

**CARMEN PAMELA ROSALES SEDANO**

**FRAMEWORK PARA IDENTIFICAÇÃO DA SEVERIDADE DE BULLYING BASEADO  
EM MACHINE LEARNING E LÓGICA FUZZY**

*Dissertação/Tese apresentada à Faculdade de  
Tecnologia da Universidade Estadual de Campinas  
como parte dos requisitos exigidos para a obtenção  
do título de Mestra em Tecnologia, na área de  
Sistemas de Informação e Comunicação.*

*Supervisor/Orientador:* Prof. Dr. Edson Luiz Ursini

*Co-supervisor/Coorientador:* Prof. Dr. Paulo Martins

ESTE EXEMPLAR CORRESPONDE A VERSÃO  
FINAL DISSERTAÇÃO/TESE DEFENDIDA PELO  
ALUNO CARMEN PAMELA ROSALES SEDANO,  
E ORIENTADA PELO(A) PROF. DR EDSON LUIZ  
URSINI

**LIMEIRA**

**2017**

**Agência(s) de fomento e nº(s) de processo(s):** Não se aplica.

Ficha catalográfica

Universidade Estadual de Campinas  
Biblioteca da Faculdade de Tecnologia  
Felipe de Souza Bueno - CRB 8/8577

R71f Rosales Sedano, Carmen Pamela, 1986  
Framework para identificação da severidade de bullying baseado em machine learning e lógica fuzzy / Carmen Pamela Rosales Sedano. – Limeira, SP : [s.n.], 2017.

Orientador: Edson Luiz Ursini.

Coorientador: Paulo Sérgio Martins Pedro.

Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Tecnologia.

1. Bullying. 2. Mineração de dados (Computação). 3. Aprendizado de máquina. 4. Máquina de vetores de suporte. 5. Lógica fuzzy. I. Ursini, Edson Luiz, 1951-. II. Martins Pedro, Paulo Sérgio, 1967-. III. Universidade Estadual de Campinas. Faculdade de Tecnologia. IV. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** A bullying-severity identifier framework based on machine learning and fuzzy logic

**Palavras-chave em inglês:**

Bullying Data mining

Machine learning

Support vectors Machine

Fuzzy logic

**Área de concentração:** Sistemas de Informação e Comunicação

**Titulação:** Mestra em Tecnologia, na área de Sistemas de Informação e Comunicação

Banca examinadora:

Edson Luiz Ursini [Orientador]

Leonardo Lorenzo Bravo Roger

Eric Alberto de Mello Fagotto

Data de defesa: 28-08-2017

Programa de Pós-Graduação: Tecnologia

## FOLHA DE APROVAÇÃO

Abaixo se apresentam os membros da comissão julgadora da sessão pública de defesa de dissertação para o título de Mestra em Tecnologia, na área de Sistemas de Informação e Comunicação, com o trabalho “FRAMEWORK PARA IDENTIFICAÇÃO DA SEVERIDADE DE BULLYING BASEADO EM *MACHINE LEARNING* E *LÓGICA FUZZY*” que submeteu a candidata Carmen Pamela Rosales Sedano, no dia 28 de agosto de 2017 na Faculdade de Tecnologia - FT/UNICAMP, em Limeira/SP.

**Prof. Dr. Edson Luiz Ursini**

Presidente da Comissão Examinadora

**Prof. Dr. Leonardo Lorenzo Bravo Roger**

FT-UNICAMP

**Dr. Eric Alberto de Mello Fagotto**

PUC-Campinas

A ata de defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica da candidata na Universidade.

## DEDICATÓRIA

*Eu quero dedicar esta dissertação a meu querido avó Dionisio que está no céu e protege a toda nossa família desde lá, aos meus pais Víctor e Carmen por seu amor infinito e confiança e especialmente à meu filho Matías que é minha maior motivação e força para ser melhor cada dia.*

## **AGRADECIMENTOS**

*Primeiramente, gostaria agradecer a Deus por ter permitido chegar a este dia porque só com Ele o impossível se torna possível. Quero expressar todo meu agradecimento ao meu orientador o Prof. Dr. Edson Luiz Ursini por acreditar em mim em todo momento, por sua compreensão, paciência e orientação durante o desenvolvimento do meu projeto.*

*Agradeço a meus avós, pais, irmãos e meu filho Matías pelo amor infinito, apoio e confiança, por acreditar em meus sonhos e em minhas capacidades profissionais e pessoais, porque eles me ensinaram a nunca desistir ainda que as situações se tornarem difíceis.*

*Agradeço também aos meus amigos do Peru por seus conselhos e confiança, aos meus amigos da Colômbia e do Brasil que conheci durante o tempo que morei no Brasil e compartilharam comigo e o Matías inesquecíveis experiências. Todos eles me demonstraram sua amizade sincera e incondicional.*

*Finalmente, agradeço especialmente ao Prof. Dr. Paulo Sérgio Martins por sua coorientação, ao Prof. Dr. Ivan de Oliveira por seu trabalho comprometido com os alunos da pós-graduação da FT, ao Prof. Dr. Ivan Marques Ricarte por sua confiança para ser PED de sua disciplina no semestre 2016-2, e a todos os professores e ao pessoal da secretaria da Faculdade de Tecnologia (FT) da UNICAMP assim como ao pessoal do Serviço de Apoio ao Estudante (SAE) por que aprendi muito deles e me apoiaram quando precisei de ajuda para continuar estudando no Brasil.*

## RESUMO

O *bullying* nas escolas é um fenômeno social sério que se apresenta em todas partes do mundo e afeta crianças e adolescentes negativamente. Contudo, os programas anti-*bullying* das escolas não deveriam se focar em rotular os estudantes como agressores ou vítimas, papéis tradicionais dos envolvidos em um episódio de *bullying*, porque aquilo produz efeitos contrários.

Portanto, é necessário uma nova abordagem que permita lidar com os episódios de *bullying*, sem precisar de rótulos mas sim de determinar o nível de severidade, assim o pessoal da escola poderá responder a aqueles episódios apropriadamente.

Os trabalhos disponíveis na literatura sobre técnicas computacionais para combater o *bullying* demonstraram resultados promissores, contudo a maioria deles oferecem informação categórica como um conjunto de rótulos. O presente projeto propõe o desenvolvimento de um framework para determinar o nível de severidade das experiências de *bullying* narradas em textos de máximo 140 caracteres. Este *framework* está composto por duas partes: (1) avaliação dos textos utilizando classificadores de *Support Vector Machine (SVM)* desenvolvidos na literatura e (2) desenvolvimento do sistema de Lógica Fuzzy, as regras deste sistema foram definidos de acordo com a literatura do *bullying* pelos autores deste projeto.

Os resultados demonstraram que é necessário melhorar a acurácia e precisão dos classificadores SVM para conseguir determinar o nível de severidade por meio do sistema de Lógica *Fuzzy*. Neste trabalho, como parte das melhorias dos classificadores SVM, se rotularam novos textos para serem utilizados na fase de pré-processamento dos dados para criar novos modelos SVM e compará-los com aqueles desenvolvidos na literatura, os quais, finalmente foram utilizados em nosso *framework*.

**Palavras-chave:** Bullying, Mineração de Textos, Aprendizado de Máquinas, Vetores de Suporte, Sistemas de Lógica Fuzzy

## ABSTRACT

Bullying at schools is a serious social phenomenon around the world that affects the development of children negatively. However, anti-bullying programs should not focus on labeling children as bullies or victims since they could produce opposite effects.

Thus, an approach to deal with bullying episodes, without labeling children, is to determine their severity, so school staff could respond them appropriately. Related works about computational techniques to fight against of bullying showed promising results but they offer categorical information as a set of labels. This work proposes a tool to determine bullying severity in texts, composed by two parts: (1) evaluation of texts using Support Vector Machine (SVM) classifiers found in literature (2) development of a Fuzzy Logic System that uses SVM classifiers outputs as its inputs to identify the bullying severity.

Results show that it is necessary to improve SVM classifiers accuracy to determine bullying severity through Fuzzy Logic. As part of our work, new texts were labeled in the data-preprocessing phase in order to develop new SVM models, which were compared to those SVM classifiers found in literature.

**Keywords:** Bullying, Text Mining, Machine Learning, Support Vector Machine, Fuzzy Logic System



## LISTA DE ILUSTRAÇÕES

Figura 1.1 – Cronograma do projeto de mestrado.....	20
Figura 2.1 – Tipos de <i>bullying</i> .....	25
Figura 2.2 – Papéis dos envolvidos em um episódio de bullying (XU; ZHU; BELLMORE, 2012).....	28
Figura 3.1 – O processo de <i>Supervised Machine Learning</i> .....	36
Figura 3.2– <i>SVM Feature Space</i> (ROSSI et al., 2015).....	37
Figura 3.3 – Hiperplano separador e Vetores de Suporte .....	37
Figura 3.4 – Exemplo do processo de <i>Tokenization</i> .....	42
Figura 3.5– Entorno de Classificação do <i>software</i> WEKA 3.2.1.....	44
Figura 3.6– Matriz de confusão para as classes “Sim” e “Não” .....	46
Figura 3.7– Conjunto <i>Crisp</i> vs Conjunto <i>Fuzzy</i> .....	48
Figura 3.8– Exemplo simples de função de pertinência .....	49
Figura 3.9– Sistema de inferência Fuzzy – (SIVANANDAM; SUMATHI; DEEPA, 2007) .....	50
Figura 3.10– Funções de Pertinência e Regras <i>Fuzzy</i> do problema da Gorjeta .....	51
Figura 3.11– Combinando saídas das regras em uma distribuição de saída .....	52
Figura 3.12– Exemplo de Sistema de Inferência <i>Fuzzy</i> com método <i>Mamdani</i> .....	53
Figura 4.1– Exemplos de tweets sobre episódios de bullying .....	58
Figura 4.2– Tarefas para o desenvolvimento do <i>framework</i> para determinar o nível de severidade em textos em inglês .....	60
Figura 4.3– Lista de <i>keywords</i> associadas ao contexto de <i>bullying</i> de acordo com a literatura .....	61
Figura 4.4– (a) Representação de texto em vetor com todos os <i>features</i> . (b) Representação de texto só com os <i>features</i> utilizados .....	61
Figura 4.5 <i>Features</i> ou índices do exemplo .....	62
Figura 4.6– Vetor normalizado para o exemplo.....	63
Figura 4.7– Número de amostras de treinamento para o Classificador “ <i>Author Role</i> ” .....	64
Figura 4.8– Número de amostras de treinamento para o Classificador “ <i>Bullying Form</i> ” .....	65
Figura 4.9– Arquivo *,arff com amostras de treinamento para o Classificador “ <i>Bullying Trace</i> ” .....	66

Figura 4.10 Fluxo para a criação de modelo SVM no <i>software WEKA</i> .....	66
Figura 4.11 Exemplo de avaliação de um texto utilizando Lógica <i>Fuzzy</i> .....	72
Figura 4.12 Rede Social " <i>Class21</i> " .....	73
Figura 4.13 Software em Java para execução do <i>framework</i> .....	74
Figura 4.14 Classificação dos <i>tweets</i> por meio do <i>framework</i> .....	74
Figura A5.1 - Determinação do potencial de "bullying".....	95

## LISTA DE TABELAS

Tabela 2.1: Programas anti-bullying com efeitos positivos em perpetração de bullying e vitimização (EVANS; FRASER; COTTER, 2014) .....	31
Tabela 2.2 - Bullying Assessment Matrix (MELHUIISH; PALMER; PASLEY, 2015) ..	31
Tabela 3.1 – Exemplos de instâncias de plantas da base de dados Iris .....	43
Tabela 3.2 – Níveis de concordância utilizando o <i>kappa</i> (LANDIS; KOCH, 1977) ....	45
Tabela 4.1 – Frequência dos <i>features</i> ou índices do exemplo .....	62
Tabela 4.2 – Descobertas das amostras de treinamento rotuladas como “não” para o Classificador “ <i>Bullying Trace</i> ” .....	64
Tabela 4.3 – Descobertas das amostras de treinamento rotuladas como “vítima” para o Classificador “ <i>Author Role</i> ” .....	65
Tabela 3.3 - Descrição dos classificadores SVM desenvolvidos em (XU, 2015).....	67
Tabela 3.4 – Acurácia dos classificadores SVM desenvolvidos em (XU, 2015).....	68
Tabela 4.4 – Regras para as variáveis de “Victim” e “General” .....	71
Tabela 4.5 – Exemplo de Tweets avaliados no <i>framework</i> .....	75
Tabela A2.1– Características dos respondentes do questionário FIPE .....	87
Tabela A3.1 – Indicadores de acordo às respostas do questionário FIPE .....	88
Tabela A3.2 – Exemplos de frases que expressa preconceito .....	88
Tabela A4.1 Atitude Preconceituosa – IPC % - para Alunos de Ensino Fundamental .....	90
Tabela A4.2 Distância Social – IPCD % - para Alunos de Ensino Fundamental .....	90
Tabela A4.3 Conhecimento de Situações de Discriminação – IPCSB % - para Alunos de Ensino Fundamental .....	91
Tabela A4.4 Conhecimento de Situações de Discriminação – IPCSB % - para alunos até 14 anos.....	91
Tabela A5.1 - Percentual das variáveis observadas nos grupos de vítimas. ....	92
Tabela A5.2 - Variáveis do modelo. ....	93
Tabela A5.3 - Percentual das variáveis na ocorrência de “bullying”. ....	94
Tabela A5.4 - “Entrevistas” obtidas por simulação. ....	94
Tabela A5.5: Valores obtidos (EY) e V_resp .....	96
Tabela B1.1 – Parâmetros do filtro “StringToWordVector” .....	97
Tabela B1.2 – Parâmetros para SVM-LIBLINEAR .....	98

Tabela B1.3 – Parâmetros para SVM-LibSVM.....	99
---	----

## LISTA DE ABREVIATURAS E SIGLAS

<b>ML</b>	–	<i>Machine Learning</i>
<b>NLP</b>	–	<i>Natural Language Processing</i>
<b>FLS</b>	–	<i>Fuzzy Logic System</i>
<b>FCL</b>	–	<i>Fuzzy Control Language</i>
<b>FIS</b>	–	<i>Fuzzy Inference System</i>
<b>SVM</b>	–	<i>Support Vector Machine</i>

# SUMÁRIO

<b>INTRODUÇÃO .....</b>	<b>13</b>
1.1 Contexto .....	13
1.2 Motivação e Objetivos .....	16
1.2.1 Objetivo Geral .....	17
1.2.2 Objetivo Específicos .....	18
1.3 Metodologia de Desenvolvimento do Trabalho .....	18
1.4 Organização do Trabalho .....	20
<b>LITERATURA SOBRE O <i>BULLYING</i>.....</b>	<b>22</b>
2.1 Definição do <i>Bullying</i> .....	22
2.2 Tipos de <i>Bullying</i> .....	24
2.2.1 <i>Bullying</i> Físico .....	24
2.2.2 <i>Bullying</i> Verbal .....	25
2.2.3 <i>Bullying</i> Social .....	25
2.2.4 <i>Bullying</i> Sexual.....	26
2.2.5 Dano à propriedade.....	26
2.2.6 <i>Cyberbullying</i> .....	26
2.3 Avaliação do nível de severidade de acordo com o tipo de <i>bullying</i> .....	27
2.4 Características dos Envolvidos em um episódio de <i>Bullying</i> .....	27
2.4.1 Padrão Vítima.....	28
2.4.2 Padrão Agressor .....	29
2.5 Programas anti- <i>bullying</i> .....	30
2.6 Considerações Finais .....	32
<b>ESTADO DA ARTE SOBRE O <i>BULLYING</i> NA CIÊNCIA DA COMPUTAÇÃO .....</b>	<b>33</b>
3.1 Revisão de conceitos de <i>Machine Learning</i> e Lógica <i>Fuzzy</i> .....	34
3.1.1 Aprendizado de Máquinas ou <i>Machine Learning</i> .....	34
3.1.2 Máquina de vetores de suporte ou <i>Support Vector Machine</i> (SVM).....	36
3.1.3 Natural Language Processing (NLP).....	39
3.1.4 Pré-processamento .....	40
3.1.5 Representação de dados .....	43

3.1.6 Medidas úteis do <i>software WEKA</i> .....	44
3.1.7 Teoria sobre Lógica <i>Fuzzy</i> ou <i>Fuzzy Logic</i> .....	47
3.2 Revisão de trabalhos.....	53
<b>PROPOSTA DE CONCLUSÃO .....</b>	<b>58</b>
5.1 Pré-Processamento dos dados .....	60
5.1.1 Tokenization .....	61
5.1.2 Representação dos <i>tweets</i> em formato vetor .....	61
5.2 Criação do modelo .....	63
5.3 Execução dos algoritmos SVM.....	67
5.4 Fuzzy Logic System para Bullying.....	69
5.4.1 Identificação da estrutura .....	69
5.4.2 Identificação de parâmetros .....	71
5.5 Desenvolvimento da rede social interna “Class21” e o Sistema em <i>Java Swing Application</i> para a aplicação do <i>framework</i> .....	73
5.6 Resultados .....	74
5.7 Conclusões.....	75
5.8 Trabalhos Futuros .....	76
<b>APÊNDICE A- MODELO ESTATÍSTICO DO TIPO LOGIT PARA AVALIAÇÃO DE POTENCIAL DE <i>BULLYING</i> NO ENSINO FUNDAMENTAL .....</b>	<b>86</b>
A1. Introdução .....	86
A2. Objetivo.....	86
A3. Natureza e Instrumentos da pesquisa da FIPE.....	87
A.4 Análise da Atitude Preconceituosa, Distância Social e Conhecimento de Situações de Discriminação ( <i>bullying</i> ) presenciadas para o público alvo .....	89
A4.1 Atitude Preconceituosa – IPC (%).....	89
A4.2 Distância Social – IPCD (%) .....	90
A4.3 Conhecimento de situações de discriminação – IPCSB (%).....	90
A5. Modelo Estatístico do tipo Logit .....	92
A5.1. Dados Obtidos inicialmente e variáveis do modelo.....	92
A5.2. Simulação a partir dos dados iniciais .....	93
A5.3 Conclusões .....	96

<b>APÊNDICE B - RESULTADOS DA FASE DE TREINAMENTO DOS ALGORITMOS PARA A CRIAÇÃO DE NOVOS MODELOS</b> .....	<b>97</b>
B1. Pré-processamento.....	97
B2. Criação do modelo “ <i>Bullying trace</i> ” .....	98
B2.1 Resultados da execução com LIBLINEAR .....	98
B2.2 Resultados da execução com LibSVM.....	99
B2.3 Resultados da execução com J48 .....	99
B3. Criação do modelo “ <i>Bullying Form</i> ”.....	100
B3.1 Resultados da execução com LIBLINEAR.....	100
B3.2 Resultados da execução com LIBSVM.....	101
B3.3 Resultados da execução com J48 .....	101
B4. Criação do modelo “ <i>Author Role</i> ” .....	102
B4.1 Resultados da execução com LIBLINEAR.....	102
B4.2 Resultados da execução com LIBSVM.....	103
B4.3 Resultados da execução com J48 .....	103
<b>APÊNDICE C – ARTIGO: APLICAÇÃO DO MIT APP INVENTOR COMO FERRAMENTA DE APOIO À APRENDIZAGEM</b> .....	<b>105</b>
<b>APÊNDICE D – ARTIGO: A SEVERITY BULLYING EPISODE IDENTIFIER FRAMEWORK BASED ON MACHINE LEARNING TECHNIQUES AND FUZZY LOGIC</b> .....	<b>112</b>



# Capítulo 1

## INTRODUÇÃO

---

---

### 1.1 Contexto

O *bullying* é um sério problema social que atinge principalmente crianças e adolescentes de Ensino Fundamental na faixa etária de 11 a 14 anos (BAUER; LOZANO; RIVARA, 2007; TUSINSKI, 2008) e pode ser encontrado em qualquer escola do mundo é por isso que diversos pesquisadores têm direcionado seus estudos para esse fenômeno que tem aspectos preocupantes tanto pelo seu crescimento, por atingir faixas etárias cada vez mais baixas. (ANTONIO et al., 2003) e consequências negativas na saúde mental e emocional dos estudantes (BARTON, 2006).

A definição do *bullying* usada neste projeto foi feita por diferentes pesquisadores entre eles Dan Olweus, que é o pioneiro dos estudos sistematizados sobre o *bullying*. O *bullying* é definido como um conjunto de atitudes agressivas para com outras pessoas e possui três características: (1) agressão intencional que pode ser física, emocional, sexual, verbal ou com o uso da tecnologia como os telefones celulares ou computadores (*cyberbullying*), (2) as vítimas estão expostas às agressões de forma repetitiva e em um período de tempo estendido e (3) apresenta-se nas relações interpessoais baseadas em um desequilíbrio de poder. No futuro, o *bullying* traz consequências negativas à saúde mental, social e emocional dos envolvidos (OLWEUS, 2003).

O ambiente inteiro da escola torna-se perigoso quando não existem intervenções efetivas contra os episódios de *bullying* e afeta a todos os alunos sem exceção (LOPES NETO, 2005). Contudo, os participantes de um episódio de *bullying*

tem papéis bem definidos. Os papéis tradicionais são o agressor (ou *bully*), a vítima (aquele que sofre as agressões) e as testemunhas ou “*bystanders*” (aqueles que só vêem o evento, mas não participam diretamente) (BARTON, 2006). Esses participantes apresentam características diferentes, no caso dos agressores, eles estão mais predispostos a estarem envolvidos em situações criminais e terem comportamentos impróprios. Por outro lado, as vítimas normalmente sofrem de depressão, solidão, ansiedade, medo e baixa autoestima (DAKE, JOSEPH A; PRICE, JAMES H; TELLJOHANN, 2003). Finalmente, as testemunhas poderiam encorajar ou inibir os episódios de *bullying* conforme com a sua empatia seja com o agressor ou a vítima respectivamente. (VAN NOORDEN et al., 2014).

De fato, os estudantes sejam agressores ou vítimas têm risco de sofrer problemas de saúde e ajuste emocional na vida adulta (EVANS; FRASER; COTTER, 2014) e, no pior dos casos, até chegar ao suicídio (JAVIER; DILLON, 2013). Portanto, o *bullying* não deveria ser considerado como uma fase “normal” do desenvolvimento das crianças mas deveria ser visto como o precursor de comportamentos de violência mais graves (PIOTROWSKI; HOOT, 2008).

Os tipos tradicionais de *bullying* são físico, verbal, material ou psicológico (SHAHEEN, 2009). Além disso, o fácil acesso das crianças e os adolescentes à Internet e às tecnologias digitais permitiu a aparição do *cyberbullying* ou *bullying online* o que se tornou uma fonte de preocupação (RICHARDSON; HIU, 2016). Em (GLADDEN et al., 2014), menciona-se que o 95% dos jovens na faixa etária de 12 até 17 anos utiliza o Internet e 80% deles possui uma conta nas principais redes sociais como *Facebook* ou *Twitter*. Deste, 9% desse reportou ser vítima de *bullying* por meio de mensagens de textos e 8% por meio de correio eletrônico.

Sendo a escola o principal lugar onde acontecem os casos de *bullying*, é importante elas desenharem e implementarem programas de prevenção e intervenção para reduzir o *bullying* e a vitimização. A efetividade desses programas promoverá resultados positivos na vida dos estudantes (EVANS; FRASER; COTTER, 2014) e deve ser um trabalho conjunto entre estudantes, professores, pessoal administrativo e pais (LOPES NETO, 2005). Para um completo entendimento sobre o *bullying*, os programas a serem desenvolvidos não devem se focar em rotular os envolvidos como “agressores” ou “vítimas” (DULMUS; SOWERS; THERIOT, 2006), porque enfatizaria o papel da criança ou adolescente sem ressaltar suas capacidades e contribuiria para

gerar um clima negativo (RAGOZZINO; UTNE O'BRIEN, 2009). Deste modo, em vez de utilizar rótulos, um efetivo programa anti-*bullying* deve ser capaz de responder aos sinais de qualquer tipo de envolvimento (RETTEW; PAWLOWSKI, 2016): “*bullying*” que está focado em ajudar aos agressores, porque o comportamento deles é só uma resposta a outras causas, como exemplo, problemas familiares (TUSINSKI, 2008) e a “vitimização” que está focada em ajudar às vítimas, incluído as testemunhas porque eles muitas vezes, preferem “sofrer em silêncio” (RAGOZZINO; UTNE O'BRIEN, 2009). Os autores em (MELHUIISH; PALMER; PASLEY, 2015) a “*Bullying Assessment Matrix*” que é uma matriz para avaliar o nível de severidade de um episódio de *bullying* como um todo sem precisar rotular os envolvidos no episódio.

Comumente, os cientistas sociais que estudam o *bullying* utilizam questionários que podem produzir informação limitada devido ao custo e à potencial fadiga dos participantes (MANCILLA-CACERES et al., 2012) além que a maioria dos questionários são só respondidos uma única vez (BELLMORE et al., 2015). Por outro lado, o estudo do *bullying* pela área de ciências da computação está surgindo com resultados satisfatórios (XU; ZHU; BELLMORE, 2012). Na atualidade, a tecnologia digital é parte da vidas dos jovens e uma de suas atividades mais frequentes o uso das redes sociais como *Facebook*, *Twitter* e *Youtube*, onde eles escrevem suas opiniões ou experiências pessoais em uma linguagem informal (FERNÁNDEZ-GAVILANES et al., 2016). Portanto, aparece uma forma nova de expressão que é uma fonte potencial de informação de muito valor, por exemplo, no caso do *Twitter*, seus usuários no ano 2013 já geravam cerca de 500 milhões de *tweets* por dia. Esta nova fonte pode ser utilizada para determinar comportamentos de *bullying* (MELHUIISH; PALMER; PASLEY, 2015) porque em alguns casos os envolvidos em episódios de *bullying* escrevem nas suas redes sociais sobre essas experiências (XU, 2015).

O trabalho de (XU, 2015) demonstra que as informações escritas pelos usuários nas redes sociais podem ser uma fonte de dados valiosa para o estudo do *bullying* se foram analisadas com apropriadas técnicas de *Machine Learning* (ML) e *Natural Language Processing* (NLP).

## 1.2 Motivação e Objetivos

As crianças e os jovens merecem se desenvolver em um ambiente livre de qualquer manifestação de violência, como o *bullying*. São eles o futuro do mundo e é nosso dever protegê-los e garantir que sua infância seja a fase na qual eles consigam desenvolver valores e atitudes positivas e assim tornarem-se adultos de bem. (GOURNEAU, 2012) refere-se ao *bullying* como um comportamento aprendido. Logo, é importante que esse tipo de comportamento agressivo seja tratado nas primeiras fases de vida assim ele pode ser “desaprendido”. Portanto, as principais instituições mundiais como a Organização das Nações Unidas para a Educação a Ciência e a Cultura (UNESCO) e O Fundo das Nações Unidas para a Infância (UNICEF) tomam medidas para prevenir e combater o *bullying*.

No último relatório da UNESCO (UNESCO, 2017) calcula-se que a proporção das crianças e jovens vítimas de *bullying* nas escolas varia de 10% a 65%, dependendo do país e no caso do *cyberbullying* varia de 5% a 25%. Nesse mesmo relatório, menciona-se que 246 milhões de crianças e jovens experimentam violência nas escolas cada ano. Em (CENTERS FOR DISEASE CONTROL, 2014) menciona-se que os envolvidos em qualquer episódio de *bullying* têm mais probabilidades de reportar maiores níveis de comportamentos suicidas do que aqueles que não. De fato, não é simples reconhecer quem está envolvido em episódios de *bullying* porque segundo (RICHARDSON; HIU, 2016) e (WOLKE; LEREYA, 2015) comumente as vítimas preferem sofrer em silêncio e resistem em contar suas experiências aos demais, devido ao medo e à vergonha; e no caso dos testemunhas (RAGOZZINO; UTNE O'BRIEN, 2009), a maioria não procura ajuda dos adultos.

As estatísticas apresentadas demonstram que o *bullying* e o *cyberbullying* são problemas reais e graves com consequências irreversíveis podendo até chegar ao suicídio. Nesse contexto, o avanço da tecnologia, corretamente direcionado, poderia ajudar na detecção com antecedência dos episódios de *bullying* e assim criar um ambiente saudável necessário para o desenvolvimento positivo de nossas crianças.

### 1.2.1 Objetivo Geral

De acordo com os fatos apresentados e também com a proposta da matriz de avaliação dos episódios de *bullying* em sua totalidade desenvolvida em (MELHUIISH; PALMER; PASLEY, 2015), o presente projeto tem como objetivo geral o desenvolvimento de um framework composto por duas partes: quatro classificadores de *Support Vector Machine* (SVM) desenvolvidos nos trabalhos de (XU, 2015)(XU et al., 2012) e um sistema de Lógica Fuzzy que em conjunto permitirão às escolas determinarem o nível de severidade de uma experiência de *bullying* descrita em textos de máximo 140 caracteres na língua inglesa por estudantes na faixa etária de 11 e 14 anos.

Decidiu-se utilizar os classificadores SVM desenvolvidos no trabalho de (XU, 2015) que pertence ao grupo de pesquisa da *University of Wisconsin-Madison* porque esse grupo já vem trabalhando desde o ano 2011 para entender e combater o *bullying* por meio de modelos novos criados com algoritmos de *Machine Learning* para responder questões científicas nessa área de interesse. Como parte de nossa contribuição, esses classificadores foram testados e melhorados antes de serem integrados no nosso *framework*. Os detalhes técnicos serão apresentados nas próximas secções.

Decidiu-se utilizar ferramentas de Lógica *Fuzzy* no projeto porque elas permitem lidar com problemas do mundo real que são complexos devido ao alto nível de imprecisão e incerteza que são gerados pela percepção própria de cada pessoa. Por exemplo, os conceitos de “muitos”, “alto”, “baixo”, “jovem”, “velho”, etc., não possuem limites definidos e só dependem do grau de avaliação de cada pessoa, portanto são conhecidos como conceitos de tipo “*fuzzy*” ou “vago” (SIVANANDAM; SUMATHI; DEEPA, 2007). No caso do problema do *bullying*, não é possível dizer que um estudante é 100% agressor ou 100% vítima, ou seja estamos diante de conceitos *fuzzy*. Por isso, nossa abordagem está focada na avaliação na totalidade do episódio de *bullying*. Além do que, essa é uma forma que parece adequada para explicar o problema para os pais dos estudantes envolvidos.

### 1.2.2 Objetivo Específicos

Para garantir o correto funcionamento de nosso *framework* de identificação do nível de severidade de um episódio de *bullying*, precisam-se atingir os seguintes objetivos específicos:

- Criação dos modelos dos classificadores SVM que respondem as seguintes questões “O texto pertence a um episódio de *bullying*?”, “Qual é o rótulo do autor que escreveu o texto?” e “Que tipo de violência foi utilizada no episódio de *bullying* descrito no texto?”.
- Validação do nosso trabalho por comparação dos resultados da execução de nossos modelos SVM com os do trabalho de (XU, 2015).
- Desenvolvimento de uma rede social interna para a coleta de novos textos a serem avaliados pelo *framework*. Essa rede social poderia ser utilizada pelos estudantes de escola fundamental.
- Desenvolvimento da interface gráfica do usuário (GUI), que permitirá ao pessoal de cada escola a execução do *framework* desenvolvido

### 1.3 Metodologia de Desenvolvimento do Trabalho

Em relação às tipologias que classificam as pesquisas, estas podem ser agrupadas em três categorias: a pesquisa quanto aos objetivos, que contempla a pesquisa exploratória, descritiva e explicativa; a pesquisa quanto aos procedimentos, que aborda o estudo de caso, levantamento de dados, a pesquisa bibliográfica, documental, participante e experimental; e a pesquisa quanto à abordagem do problema, que compreende a pesquisa qualitativa e a quantitativa (BEUREN, 2015).

Quanto aos objetivos, este trabalho pode ser classificado como uma pesquisa exploratória porque no início não se tinha uma hipótese ou objetivo definido em mente e foi na análise dos resultados de trabalhos da literatura e na busca de anomalias e fatores ainda não conhecidos (WAZLAWICK, 2010) que permitiram definir os nossos

próprios objetivos e comparar nossos resultados com aqueles outros trabalhos, focados em atingir a correta detecção do nível de severidade dos episódios de *bullying* escritos em textos; é também descritiva porque no processo do desenvolvimento de nosso *framework*, os dados coletados para o treinamento e teste dos algoritmos SVM foram de fontes como *Twitter* e sítios webs utilizando palavras chaves associadas ao problema do bullying que ajudaram a entender e descrever como o público alvo narra suas experiências de *bullying* pela internet (WAZLAWICK, 2010).

Com referência aos procedimentos, a pesquisa é bibliográfica, pois emprega conceitos, teorias e proposições ao longo do texto, sobretudo no referencial teórico, por meio de livros e artigos científicos (BEUREN, 2015); é também pesquisa de levantamento, porque os dados utilizados do Twitter (*tweets*) foram coletados diretamente com ferramentas como a interface de programação de aplicações (API) do próprio Twitter disponível para os programadores, esses *tweets* foram rotulados com variáveis de interesse para os algoritmos SVM e do sistema de Lógica Fuzzy. Finalmente, é uma pesquisa experimental porque o objetivo geral do projeto tem aplicação prática em um *framework* que inclui as melhoras do código dos algoritmos SVM desenvolvidos em (XU, 2015) e a integração com nosso sistema de Lógica Fuzzy (WAZLAWICK, 2010).

Em relação à abordagem do problema, a pesquisa se classifica por um lado como uma pesquisa qualitativa, pois é a forma mais adequada de conhecer a natureza do fenômeno social, além de possibilitar descrever a complexidade do problema abordado que é o *bullying*, analisa-lo e compreende-lo (RICHARDSON, 2007); e é quantitativa, porque os textos coletados sobre experiências de bullying são convertidos em formato de vetores constituídos por valores numéricos para, em seguida, serem avaliados e então, obter um número que representa o nível de severidade.

Na Figura 1, apresentam-se as principais fases envolvidas no desenvolvimento do projeto de mestrado. As disciplinas obrigatórias e eletivas necessárias para obter o número mínimo de créditos do mestrado foram feitas nos semestres 2014-1, 2015-2 e 2016-1; durante o semestre 2015-2, se realizou uma revisão bibliográfica sobre o *bullying* com os trabalhos já realizados na literatura para combater esse problema do ponto de vista da Ciência da Computação. Assim pudemos definir o tema e os objetivos do projeto de mestrado que foram apresentados na qualificação no início do

semestre 2016-1. Nessa qualificação, a banca deu suas sugestões e comentários que permitiram focar a abordagem do tema do presente projeto. Finalmente durante o semestre 2016-1 e 2016-2 se desenvolveu o *framework* para identificação do nível de severidade em textos que incluíam as etapas de pré-processamento de dados, treinamento e teste de algoritmos SVM, definição das regras para o Sistema de Lógica *Fuzzy* e a integração entre estes dois últimos processos.

Como parte do curso de mestrado, foram apresentados dois artigos em congressos internacionais, eles estão disponíveis nos Apêndices C e D

Semestre 2014-1	Semestre 2014-2	Semestre 2015-1	Semestre 2015-2	Semestre 2016-1	Semestre 2016-2	Semestre 2017-1
Disciplinas Obrigatórias (1)			Disciplinas Obrigatórias (1)		PED C	
Disciplinas Eletivas (1)	Trancamento Matrícula	Trancamento Matrícula	Disciplinas Eletivas (2)	Disciplinas Eletivas (2)	Disciplinas Eletivas (1)	
			Exame Proficiência em Inglês			
			Definição do tema do Projeto	Desenvolvimento do <i>bullying-severity identifier framework</i>		
			Revisão Bibliográfica			
			Escrita Qualificação			Escrita e Apresentação de Artigos em congressos
				Qualificação		Melhoras aos classificadores SVM
				Escrita Dissertação		

Figura 1.1 – Cronograma do projeto de mestrado

## 1.4 Organização do Trabalho

O presente projeto está organizado da seguinte forma: o Capítulo 2 apresenta a revisão da literatura sobre o *bullying* para um melhor entendimento do problema assim como as abordagens tradicionais utilizadas para reduzir os episódios de *bullying* nas escolas; o Capítulo 3 apresenta conceitos e técnicas de Machine Learning focados na classificação de textos, conceitos de Lógica *Fuzzy* e o estado da arte sobre o combate do *bullying* como área de interesse na Ciência da Computação; no Capítulo



4, apresenta-se a metodologia utilizada para o desenvolvimento do framework para a identificação do nível de severidade em textos e os resultados após de avaliar 18,400 textos (*tweets*) de 140 caracteres coletados desde junho 2016 até junho 2017.

# Capítulo 2

## LITERATURA SOBRE O *BULLYING*

---

---

O *Bullying* é um importante problema de saúde pública que tem gerado muito interesse e atenção por parte dos cientistas e pesquisadores sociais nos últimos 40 anos (EVANS; SMOKOWSKI, 2016). Contudo, um fator crítico para o entendimento do *bullying* é distingui-lo de qualquer outro tipo de agressão entre os jovens já que o *bullying* possui características únicas (GLADDEN et al., 2014).

### 2.1 Definição do *Bullying*

Segundo (EVANS; SMOKOWSKI, 2016) o estudo do *bullying* apresenta um importante ponto fraco que é a ausência de uma definição padrão e consistente sobre o *bullying*. A definição mais utilizada nos programas anti-*bullying*, desenvolvidos pelas escolas mundialmente, foi estabelecida pelo Dr. Dan Olweus e utilizada no questionário *Olweus Bullying/Victimization*: “Um estudante está sofrendo *bullying* se ele/ela está exposto a (1) comportamentos agressivos e negativos; (2) intencionais e repetidas (3) em uma relação desigual de poder e força”. (OLWEUS, 2003)

Uma definição mais abrangente inclui o *bullying* por meio da exclusão e foi fornecida pela pesquisa do *Health Behavior in School-Aged Children*: “Podemos dizer que um estudante é vítima de *bullying* quando outro estudante ou um grupo de estudantes lhe dizem ou lhe fazem coisas desagradáveis ou indecentes. Além disso, o estudante é provocado repetitivamente de uma forma que ele/ela não gosta ou é excluído deliberadamente. Porém, não pode ser considerado *bullying* quando dois estudantes com a mesma força ou poder discutem ou brigam. Também não é *bullying*

se o estudante é provocado de forma divertida ou amigável” (IANNOTTI, 2010). Em (TUSINSKI, 2008) encontra-se uma característica adicional que é ausência de provocação por parte da vítima e em (RETTEW; PAWLOWSKI, 2016) menciona-se que a desigualdade de poder entre o agressor e a vítima está em termos de tamanho físico, status social ou outros fatores. Em (GOURNEAU, 2012) adiciona-se um quarto componente, além dos mencionados, o “terror”, que atinge fortemente as vítimas para mantê-las controladas e intimidadas.

Diversas pesquisas, relatórios e questionários (DULMUS; SOWERS; THERIOT, 2006; BAUER; LOZANO; RIVARA, 2007; VAN GOETHEM; SCHOLTE; WIERS, 2010; SAINIO et al., 2012; GOLDWEBER; WAASDORP; BRADSHAW, 2013; EVANS; FRASER; COTTER, 2014) para avaliação do *bullying* utilizam uma definição baseada nos três componentes essenciais de (OLWEUS, 1993). Em (EVANS; FRASER; COTTER, 2014) salienta-se o uso da definição do *bullying* nos programas anti-*bullying* especialmente nos questionários para os estudantes responderem com uma definição padrão e não com suas próprias interpretações.

Nesse contexto, a *Center for Disease Control (CDC)* em (GLADDEN et al., 2014) propõe a seguinte definição padrão baseada também em (OLWEUS, 1993): O *bullying* entre os jovens é qualquer comportamento(s) agressivo de um jovem ou um grupo de jovens, que não são parentes nem parceiros, que envolve uma desigualdade de poder percebido ou observado, além de se repetir muitas vezes ou que é muito provável de se repetir. O *bullying* pode causar sofrimento nos jovens alvo ou vítimas por meio de dano físico, psicológico, social ou educacional.

A definição da CDC tenta garantir ao pessoal das escolas avaliarem e discutirem o *bullying* por meio da mesma linguagem e ideias. Os programas anti-*bullying* e práticas disciplinares conseguiriam responder aos episódios de *bullying* com menos dificuldade quando utilizarem uma correta definição do que é *bullying* (GLADDEN et al., 2014).

Neste projeto de mestrado não se está considerando a característica de “repetição ou frequência” da definição do *bullying*.

## 2.2 Tipos de *Bullying*

O comportamento agressivo das pessoas envolvidas em episódios de *bullying* é complexo e pode ser de muitos tipos como agressão o contato físico, verbal, social por meio de exclusão ou criando rumores, sexual e dano a propriedade (BARTON, 2006) (OLWEUS, 2003) os quais se apresentam em um entorno físico e/ou digital, ambos podendo causar dor (MELHUIISH; PALMER; PASLEY, 2015). No caso do entorno físico, o pessoal das escolas deveriam encorajar aos professores, pais e alunos a observarem e a reagirem bem aos sinais de *bullying* (MELHUIISH; PALMER; PASLEY, 2015); mas no caso do entorno digital onde a informação é duplicada, distribuída e acessada desde diversas localizações e em qualquer momento além de deixar registros permanente (por exemplo, fotos ou textos na internet), gera maiores possibilidades para as pessoas atuarem de forma negativa e principalmente anônima, portanto dificulta aos pais e professores a ajudarem (GOURNEAU, 2012) (JAVIER; DILLON, 2013).

Esses comportamentos agressivos podem ser considerados “descobertos/diretos” e fáceis de serem observados ou “cobertos/indiretos” (MELHUIISH; PALMER; PASLEY, 2015). Como parte do estudo do *bullying* é preciso distinguir as características de cada uns dos tipos (GOLDWEBER; WAASDORP; BRADSHAW, 2013). Na Figura 2, apresenta-se de forma resumida, uma classificação dos comportamentos agressivos de acordo com o tipo e o entorno onde acontecem.

### 2.2.1 *Bullying* Físico

O *bullying* físico consiste em agressões físicas com a intenção de causar dor e desconforto aos outros (UNESCO, 2017) como por exemplo bater, chutar, socar, empurrar, entre outras ações que envolvam contato físico negativo (ANTONIO et al., 2003).

### 2.2.2 Bullying Verbal

Algumas ações que podem estar presentes no *bullying* verbal são: colocar apelidos com o fim de ferir os sentimentos, humilhar e discriminar com palavras, ameaçar e intimidar aos outros (GOURNEAU, 2012)(ANTONIO et al., 2003). Em (GOURNEAU, 2012) indica-se que os apelidos e as humilhações utilizados estão focados na aparência, raça, etnia, religião, gênero e status social das vítimas, também considera-se este tipo de *bullying* como o primeiro passo para ir em direção de comportamentos mais degradantes e cruéis.

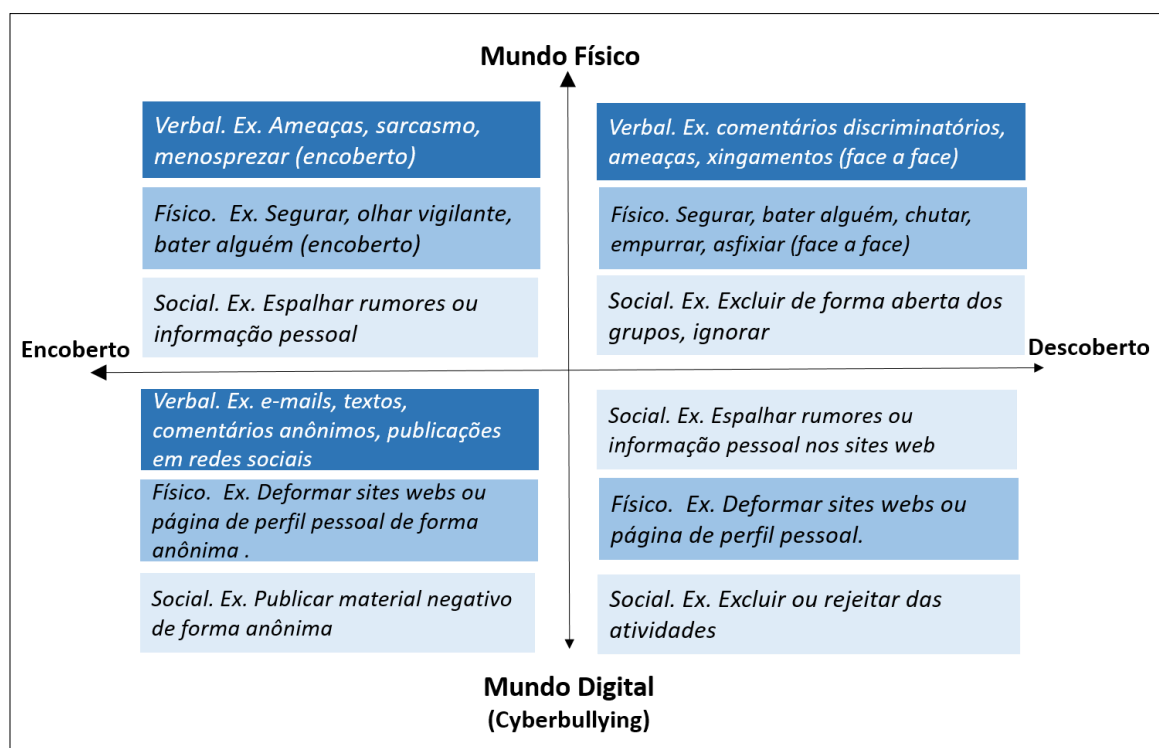


Figura 2.1 – Tipos de *bullying*

### 2.2.3 Bullying Social

O *bullying* social, algumas vezes chamado de *bullying* emocional ou psicológico (SHAHEEN, 2009), inclui o espalhamento de falsos rumores, prejudicar a reputação a exclusão intencional de alguém dos grupos ou das atividades sociais (DULMUS; SOWERS; THERIOT, 2006) e o uso das relações entre as pessoas com o fim de prejudicar os outros, este último comportamento é também conhecido como “*bullying* relacional” (CRICK; GROTPETER, 1996). Duas características principais de este tipo

de bullying é que causa dor as vítimas sem precisar de palavras ou agressões físicas; é sutil e mais difícil de detectar (RAGOZZINO; UTNE O'BRIEN, 2009) e, portanto, não gera a atenção necessária por parte dos professores e adultos (PÉREZ, 2011). A seguir, um exemplo de *bullying* social: uma estudante reportou que uma colega da turma parou de falar com ela e encorajou aos outros a fazerem o mesmo (PIOTROWSKI; HOOT, 2008).

#### **2.2.4 Bullying Sexual**

Esse tipo de *bullying* inclui muitas das ações típicas de qualquer comportamento de *bullying* mas adiciona outros componentes como exibicionismo, voyeurismo, propostas sexuais, assédio e abuso sexual (ONG, 2003).

#### **2.2.5 Dano à propriedade**

O dano à propriedade é definido como roubo, alteração ou destruição da propriedade de uma pessoa com o objetivo de causar sofrimento (EVANS; FRASER; COTTER, 2014). Esse tipo de comportamento também pode incluir a destruição não só de coisas materiais, como remover ou danificar informação digital (GLADDEN et al., 2014).

#### **2.2.6 Cyberbullying**

O *cyberbullying* é um tipo de comportamento agressivo que permite aos estudantes continuarem fazendo *bullying* fora do horário da escola, por meio de dispositivos eletrônicos como *smartphones*, *laptops*, *tablets* entre outros (BARTON, 2006) (WOLKE; LEREYA, 2015). Os estudantes que sofrem de *cyberbullying* têm que lidar com correios eletrônicos desmoralizadores e mensagens de texto (VAN NOORDEN et al., 2014) inclusive na segurança da suas próprias casas, é importante levar em conta que os agressores nesta modalidade utilizam as mesmas palavras nos textos como se fosse pessoalmente (GOURNEAU, 2012), além de aproveitarem a possibilidade de publicar e espalhar comentários/fotos obscenos e grosseiros de

forma anônima (ONG, 2003) o que dificulta aos adultos detectarem ou acompanhar os fatos (RAGOZZINO; UTNE O'BRIEN, 2009).

### **2.3 Avaliação do nível de severidade de acordo com o tipo de *bullying***

Os profissionais médicos consideram que é útil diferenciar os níveis mais baixos de *bullying* (como os de tipo verbal e social) de aqueles com alto nível (ameaças, intimidação e agressão física), embora deve-se entender que todos os tipos de *bullying* podem ser potencialmente prejudicial para os envolvidos (RETTEW; PAWLOWSKI, 2016). Em (PÉREZ, 2011) realizou-se uma pesquisa para avaliar a percepção dos professores de escola de ensino fundamental sobre a gravidade do *bullying* de tipo físico, verbal e social ou relacional e o resultado foi que o *bullying* físico é o mais grave de todos os tipos. Contudo, a forma mais comum de *bullying* é de tipo verbal (colocar apelidos, informar) seguido de *bullying* social (espalhar rumores), e o menos frequente é o *bullying* físico (JAVIER; DILLON, 2013).

É importante considerar que os agressores podem utilizar todos os tipos de *bullying* seja de forma separada ou misturada (PIOTROWSKI; HOOT, 2008) e o impacto de qualquer tipo de violência durante o desenvolvimento e bem-estar das crianças é penetrante, grave e duradouro. No caso do *bullying* e o *cyberbullying*, eles estão rodeados de um sentimento profundo de medo, solidão e desamparo (RICHARDSON; HIU, 2016).

### **2.4 Características dos Envolvidos em um episódio de *Bullying***

Comumente, os agressores e as vítimas são os participantes principais em um episódio de *bullying* ocorrido na escola (OLWEUS, 2003). Entretanto, o *bullying* não é um evento isolado entre só dois indivíduos (BARTON, 2006). De fato, outros estudantes jogam papéis bem definidos e possuem diferentes atitudes ou reações diante uma situação de *bullying* (OLWEUS, 2003). Em (BARTON, 2006), apresenta-

se a relação “Agressor – Vítima - Testemunha”, além disso divide o papel de testemunha em três tipos: não envolvidos (porque só assistem ao que está acontecendo), ajudantes do agressor (estudantes que estimulam ao agressor sem entrar em ação contra a vítima) e defensores da vítima (estudantes que intervêm em nome da vítima durante o episódio de *bullying*). Adicionalmente, o trabalho de (XU; ZHU; BELLMORE, 2012) menciona dos novos papéis para o caso das redes sociais: “repórter” que é alguém que poderia não estar presente durante o episódio mas relata o episódio e “acusador” que culpa ou acusa alguém como agressor.

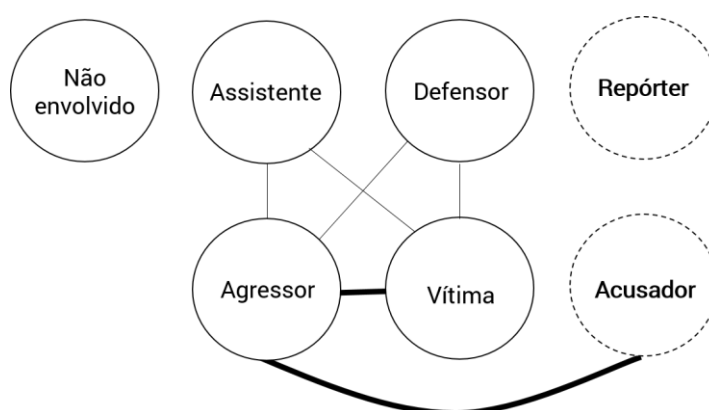


Figura 2.2 – Papéis dos envolvidos em um episódio de bullying (XU; ZHU; BELLMORE, 2012)

### 2.4.1 Padrão Vítima

Em (RICHARDSON; HIU, 2016)(WOLKE; LEREYA, 2015) menciona-se que todas as crianças estão em risco de serem vítimas de bullying, mas algumas delas devido a suas situações específicas, têm maior predisposição a sofrerem de estigmatização, discriminação ou exclusão. Por exemplo incluem-se as crianças com deficiências; refugiados; que pertencem a um determinado grupo étnico, racial, social, cultural ou religioso; também por sua orientação sexual ou identidade de gênero (MELHUIISH; PALMER; PASLEY, 2015).

Em (TUSINSKI, 2008) (ANTONIO et al., 2003) apresentam-se alguns sinais que poderiam indicar que uma criança está sendo vítima de *bullying* como: resistência em ir à escola, mau rendimento escolar, chegar muitas vezes em casa com machucados inexplicáveis, pedir para trocar de escola ou simplesmente não tem interesse na escola. É preciso saber como interpretar esses sinais porque comumente as crianças



vítimas de *bullying* são retraídas ou fechadas e portanto têm dificuldade para falar sobre o que está acontecendo e pedir ajuda.

Apesar de que os efeitos de qualquer tipo de *bullying* poderem diferir de uma vítima para outra, em essência, eles violentam a integridade e dignidade das crianças como exemplos desses efeitos são: ansiedade, depressão, medo, angústia, confusão, raiva, insegurança, baixa autoestima, profunda sensação de exclusão, impotência e abandono (RICHARDSON; HIU, 2016). O bullying produz em alguns casos distúrbios somáticos (cefaleia, desmaios, vômitos, dores em extremidades, paralisias, queixas visuais, enurese noturna, alterações do sono) (EVANS; FRASER; COTTER, 2014), distúrbios alimentícios (anorexia, bulimia) e distúrbios psiquiátricos e no pior dos casos chegar até a idealização e tentativa de suicídio como forma de buscar a solução dos problemas (DAKE, JOSEPH A; PRICE, JAMES H; TELLJOHANN, 2003) (GOURNEAU, 2012).

#### **2.4.2 Padrão Agressor**

Os agressores com frequência possuem tendências psicológicas complexas e contraditórias o que contribui a terem altos níveis de dor emocional que podem tomar diferentes formas. Por exemplo, alguns agressores têm persistente sentimentos de inferioridade, insegurança, inadequação o que termina em auto ódio. Essas crianças preferem se tornar em agressores para mitigar seus sentimentos de frustração e fracasso. Outros parecem que têm alta autoestima e utilizam comportamentos de *bullying* para consolidar seus sentimentos de valor próprio. Tem outro grupo de agressores que reportam não gostar de se comportar assim mas não tem as habilidades para mudar seus comportamentos (PIOTROWSKI; HOOT, 2008).

Em muitos trabalhos (GOURNEAU, 2012)(TUSINSKI, 2008)(VAN NOORDEN et al., 2014) menciona-se que os agressores possuem pouca maturidade emocional o que limita sua capacidade de ter empatia com os outros. Essa falta de empatia permite aos agressores utilizarem qualquer tipo de *bullying* sem sentir arrependimento pelas vítimas e justificarem suas ações violentas.

Em (DAKE, JOSEPH A; PRICE, JAMES H; TELLJOHANN, 2003)(TUSINSKI, 2008) explicam que existem muitos fatores que possibilitam a formação da

personalidade agressiva: existência de um ambiente familiar adverso e hostil, relacionamento afetivo pobre entre os membros da família, prática de maus tratos físicos e abuso (exemplo: os pais usam castigos para disciplinar), os pais que podem ser muito permissivos ou autoritários. De acordo a (ONG, 2003) a mídia (rádio, televisão, internet) também poderia influir negativamente já que às vezes fortalecem a ideia que o sucesso só é atingido sendo agressivo.

## 2.5 Programas anti-*bullying*

Muitos cientistas enfatizam a necessidade de desenvolver novos métodos e técnicas para lidar com o fenômeno do *bullying* e empoderar os estudantes e professores e abordá-lo eficientemente (RICHARDSON; HIU, 2016).

Na revisão sistemática de (EVANS; FRASER; COTTER, 2014) avaliou-se testes de programas anti-*bullying* publicados desde Junho, 2009 até Abril, 2013; os critérios de inclusão e exclusão utilizados nessa revisão permitiram encontrar 32 artigos que avaliaram 24 programas diferentes; cada artigo tinha um objetivo classificado em (a) perpetração de *bullying* e vitimização (17 artigos), (b) vitimização unicamente (10 artigos) e (c) perpetração de *bullying* (5 artigos). A maioria desses artigos coletou informação por meio de questionários dos estudantes, relatórios dos professores e observações. Além disso, na revisão se definiu que para um programa ser considerado como “medido rigorosamente”, o programa deveria medir os diferentes tipos de comportamento de *bullying* (GLADDEN et al., 2014) e oferecer uma definição abrangente do *bullying*. Na Tabela 2.1, se apresentam os programas avaliados em (EVANS; FRASER; COTTER, 2014) com maior impacto positivo em mitigar a perpetração do *bullying* e vitimização nas escolas.

Entre os programas com resultados positivos da Tabela 2.1, unicamente o programa “*Kiva*”, que foi realizada na Finlândia com um amostra homogênea, reportou efeitos significados e atingiu os critérios de rigorosidade. De fato, a maioria dos estudos com maior sucesso na prevenção do *bullying* foram realizadas fora dos Estados Unidos; isto acontece porque as amostras refletem a grande heterogeneidade que lá existe. Deste modo, estas descobertas sugerem que abordagens de prevenção

do bullying focada no aspecto cultural poderiam ser mais efetivas do que aquelas desenvolvidas para uma população genérica. Outra evidencia é que programas criativos e não tradicionais que incluem atividades em computador como jogos interativos, parecem levar a maiores mudanças nos comportamentos de *bullying*. (EVANS; FRASER; COTTER, 2014).

Tabela 2.1: Programas anti-bullying com efeitos positivos em perpetração de bullying e vitimização (EVANS; FRASER; COTTER, 2014)

Programa anti-bullying	Mudança em Perpetração do bullying	Mudança em Vitimização	Tem definição de bullying?	Possui atividades em computador ou vídeos
<i>Cool Kids</i>	Não	Sim	Não	Não
<i>Empathy Training</i>	Sim	Não	Não	Não
<i>FearNot!</i>	Sim	Sim	Não	Sim
<i>KiVa</i>	Não	Sim	Sim	Sim
<i>Olweus Bullying Prevention</i>	Misto	Misto	Sim	Sim
<i>Positive Action</i>	Sim	Não	Não	Não
<i>Take the LEAD</i>	Sim	Sim	Não	Não

Em (MELHUIH; PALMER; PASLEY, 2015) considera-se que todos os episódios de *bullying* têm que ser respondidos com políticas e processos de acordo a suas naturezas. Portanto propõe a “*Bullying Assessment Matrix*” (matriz de avaliação do *bullying*) para avaliar os três domínios de um episódio de *bullying*: severidade, impacto e frequência de acordo com três níveis (moderado, importante, severo). Essa matriz ajudaria as escolas na tomada de decisões sobre como reagir diante de um episódio.

Tabela 2.2 - Bullying Assessment Matrix (MELHUIH; PALMER; PASLEY, 2015)

	Nível do episódio de <i>bullying</i>		
	Moderado Score: 1	Importante Score: 2	Severo Score: 3
<b>Severidade</b>	Ameaças físicas ou danos, intimidação e exclusões sociais.	Ameaças físicas ou danos, intimidação, declarações ou ameaças sexuais.	Danos físicos que precisam atenção médica, ameaças sexuais, declarações que incitam ao suicídio.
<b>Impacto</b>	As vítimas superam bem e precisam só de apoio mínimo.	As vítimas superam bem e precisam de um período adicional de apoio na escola.	As vítimas são vulneráveis ou precisam de apoio intensivo da escola ou de especialistas.
<b>Frequência</b>	Nunca ou raramente aconteceu antes e é muito improvável que seja replicado digitalmente.	Episódios similares já aconteceram menos de três vezes ou é provável que seja replicado digitalmente.	Episódios similares já aconteceram mais de três vezes ou é muito provável que seja replicado digitalmente.

Na tabela 2.2, apresentam-se as características em cada combinação de domínio e nível do episódio. A avaliação total do episódio é assoma do escore atribuído a cada dimensão. Desta maneira, o episódio de bullying é considerado moderado, importante ou severo se o escore total varia entre 8-9, 6-7 ou 3-5 respectivamente.

## 2.6 Considerações Finais

Existem diversas abordagens para lidar com o *bullying* como aquelas que se focam exclusivamente nos atores de um episódio de *bullying*, outras que buscam mudar em sua totalidade a cultura de cada escola da comunidade e assim ela seja capaz de resolver os incidentes de bullying (RICHARDSON; HIU, 2016). Contudo, antes de analisar o que pode ser feito na escola, este capítulo tentou dar todas as informações necessárias para um completo entendimento do problema e as consequências negativas que se produzem na vida dos envolvidos na violência escolar.

Um fato importante de acordo (GLADDEN et al., 2014) é que apesar que o uso da tecnologia produzir um novo contexto para o *bullying* acontecer (*cyberbullying*), as pesquisas nacionais mais importantes dos Estados Unidos como o *Youth Risk Behavior Survey (YRBS)* e *the School Crime Supplement (SCS)* conferiram que uma maior porcentagem de estudantes sofrem de bullying “em pessoa” mais do que no mundo digital ou *cyberbullying*. Este aspecto foi levado em conta em nosso projeto para demonstrar a necessidade não só de detectar casos de *cyberbullying* mas tentar identificar outros tipos de bullying que podem ser encontrados quando os estudantes relatam suas experiências nas redes sociais.

# Capítulo 3

## ESTADO DA ARTE SOBRE O *BULLYING* NA CIÊNCIA DA COMPUTAÇÃO

---

A maioria de pesquisas para combater o *bullying*, principalmente do ponto de vista da Ciência da Computação, estão focados na identificação automática do *cyberbullying* ou expressões de ódio e logo reportá-los ou bloqueá-los, sem considerar a presença de *bullying* no mundo físico ou real como é feito do ponto de vista das Ciências Psicológicas (BELLMORE et al., 2015) (XU, 2015). Contudo, novas fontes de dados aparecem para enriquecer o estudo do *bullying*: as redes sociais; elas são de grandes proporções, dinâmicas e disponíveis em tempo real além de serem um elemento importante no contexto social dos jovens (WANG; IANNOTTI; NANSEL, 2009).

O *bullying* está presente nas redes sociais não só em forma de *cyberbullying*, em (XU; ZHU; BELLMORE, 2012; BELLMORE et al., 2015; XU, 2015) afirma-se que os participantes de um episódio de *bullying* poderiam publicar textos sobre suas experiências, valiosos fragmentos chamados de “rastros de *bullying*” (*bullying trace*) que inclusive vão além do *cyberbullying* porque esses rastros são respostas *online* dos tipos de *bullying* tradicional (verbal, físico, social). O presente projeto continua a pesquisa dos trabalhos de (XU; ZHU; BELLMORE, 2012; BELLMORE et al., 2015; XU, 2015) sobre o uso de técnicas de *Machine Learning* para entender o *bullying* por meio das redes sociais e adiciona regras por médio de Lógica *Fuzzy* para uma avaliação mais completa principalmente com relação à severidade dos episódios ocorridos.

A seguir, se explicam os conceitos principais utilizados na proposta do projeto e os projetos de pesquisa no estado da arte sobre o combate do *bullying* do ponto de vista da Ciência da Computação.

### **3.1 Revisão de conceitos de *Machine Learning* e Lógica Fuzzy**

#### **3.1.1 Aprendizado de Máquinas ou *Machine Learning***

O termo *Machine Learning* foi cunhado por Arthur Lee Samuel e significava a inclusão de muitas atividades inteligentes que poderiam ser transferidas desde um humano para uma máquina. O conceito de máquina deve ser entendido de forma abstrata: não como um dispositivo físico mas como um sistema automatizado, por exemplo um software (GUYON; ELISSEEFF, 2006).

Os problemas de *Machine Learning* acontecem quando uma atividade é definida por uma série de casos ou exemplos em vez de regras pré-definidas. Esses problemas podem ser encontrados em uma ampla variedade de domínios, desde aplicações de engenharias em robótica e reconhecimento de padrões (fala, manuscrito, reconhecimento facial) até aplicações na Internet (classificação de textos) e aplicações médicas (diagnóstico, prognóstico, descoberta de drogas) (GUYON; ELISSEEFF, 2006).

Cada instância de qualquer *dataset* ou conjunto de dados utilizado pelos algoritmos de *Machine Learning* é representada com o mesmo conjunto de *features* ou características. Essas *features* podem ser contínuas, categóricas ou binárias. Se as instâncias estão associadas a um rótulo ou classe definido manualmente, ou seja se já possui a resposta correta com antecedência, então estamos diante de um aprendizado supervisionado (*Supervised Learning*) (KOTSIANTIS, 2007) em contraste com o aprendizado não-supervisionado (*Unsupervised Learning* ou *Clustering*) (JAIN; MURTY; FLYNN, 1999).

No aprendizado supervisionado tem-se a figura de um professor externo, o qual apresenta o conhecimento do ambiente por conjuntos de exemplos na forma: entrada, saída desejada. O algoritmo de *Machine Learning* extrai a representação do

conhecimento a partir desses exemplos. O objetivo é que a representação gerada seja capaz de produzir saídas corretas para novas entradas não apresentadas previamente. Por enquanto, no aprendizado não-supervisionado não há a presença de um professor, ou seja, não existem exemplos rotulados; e os algoritmos aprendem a representar (ou agrupar) as entradas submetidas segundo uma medida de qualidade (LORENA; DE CARVALHO, 2007) (JAIN; MURTY; FLYNN, 1999).

O aprendizado abordado neste trabalho é o do tipo supervisionado. Portanto, apresenta-se um exemplo prático do uso de algoritmos de aprendizado supervisionado como a detecção de *spam* (BOMMERSON, 2015): Para filtrar as mensagens, o processo começa com o pré-processamento dos dados que consiste na limpeza deles para garantir sua qualidade; logo, deve-se definir o conjunto de mensagens rotulados como “não spam” e “sim spam” que são o *training set* ou conjunto de dados de treinamento que ajudará a ajustar e a selecionar os parâmetros adequados do algoritmo selecionado por meio do “treinamento” para gerar um modelo preditivo ou “classificador”. Comumente se avalia esse modelo de acordo com a sua precisão para efetuar predição (KOTSIANTIS, 2007). Finalmente, após do treinamento vem a fase de “generalização”, em que aplica-se o modelo criado sobre um conjunto desconhecido chamado de *test set* ou dados de teste para predizer seus rótulos (KOTSIANTIS, 2007). Se nesta fase os resultados são positivos, o modelo é colocado em produção. Existem muitos tipos de algoritmos supervisionado que realizam esta tarefa, contudo as diferenças entre os diversos algoritmos estão nos resultados da classificação (FISSETTE, 2010). Portanto, a escolha do algoritmo é um passo crítico (KOTSIANTIS, 2007). A Figura 3.1 mostra os passos explicados para este exemplo.

Em (KOTSIANTIS, 2007) menciona-se que a tarefa de classificação supervisionada é uma das tarefas mais frequentes realizadas pelos Sistemas Inteligentes. Os algoritmos desenvolvidos para realizar essa tarefa estão divididos em técnicas de Lógica/Símbolos como árvores de decisão (*decision trees*) e classificadores baseados em regras (*rule-based classifiers*); redes neurais; algoritmos estatísticos como classificadores *Naive Bayes* e Redes Bayesianas; algoritmos *lazy learning* como o *Nearest Neighbour*, e uma técnica moderna na área de *Machine Learning* que são as Máquinas de vetores de suporte (*Support Vector Machine* ou *SVM*) (GUGGENBERGER, 2012).

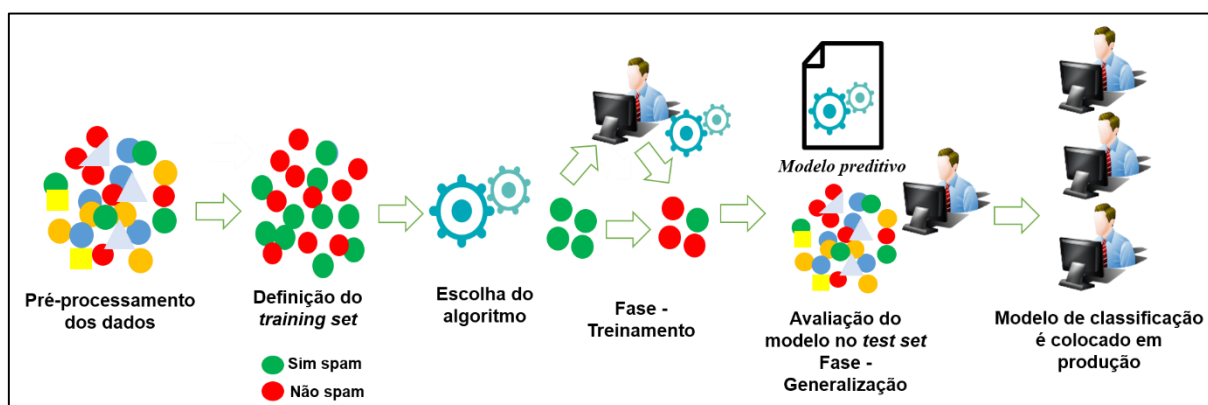


Figura 3.1 – O processo de *Supervised Machine Learning*

O presente projeto utiliza SVMs, portanto se aprofundará nos conceitos associados a esta técnica.

### 3.1.2 Máquina de vetores de suporte ou *Support Vector Machine (SVM)*

As Máquinas de Vetores de Suporte (SVM) constituem uma técnica de aprendizado que vem recebendo crescente atenção da comunidade de *Machine Learning* (LORENA; DE CARVALHO, 2007) e têm sido aplicadas amplamente em problemas de classificação (WANG, 2005).

Considerando um problema típico de classificação, alguns vetores de entrada (*feature vectors*) possuem um rótulo ou classe e o objetivo do SVM é produzir um modelo baseado nesses dados de treinamento capaz de prever a classe dos novos vetores fazendo com que a taxa de erro de classificação seja a mínima possível (HSU; CHANG; LIN, 2016) (GUGGENBERGER, 2012). Em 1965, Vladimir Vapnik introduz uma abordagem matemática para resolver esse tipo de otimização por meio da projeção dos dados de treinamento a um espaço de maior dimensão chamado de espaço de características ou "*feature space*" porque nessa nova dimensão é mais fácil separar os dados de forma linear por meio de um hiperplano (GUGGENBERGER, 2012; ROSSI et al., 2015) como se mostra na Figura 3.2.



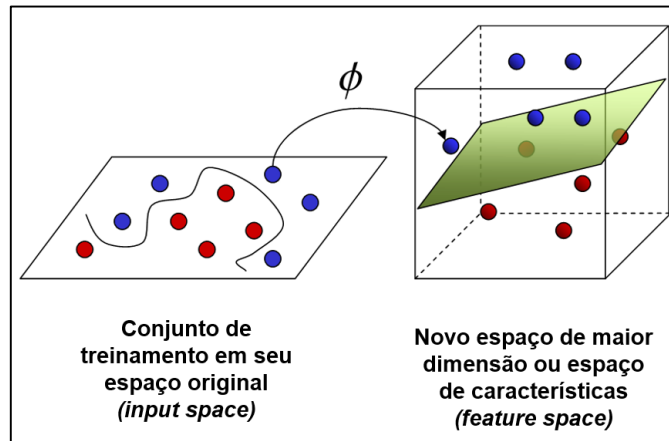


Figura 3.2– SVM Feature Space (ROSSI et al., 2015)

Os SVMs são chamados também de classificadores de máxima margem devido a que o hiperplano resultante deve maximizar a distância entre os vetores “mais próximos” das diferentes classes (positiva e negativa), com a suposição que uma margem maior permitirá minimizar o erro durante a “generalização”. Esses vetores “mais próximos” são chamados de vetores de suporte e só eles serão considerados para a tarefa de classificação, os demais serão ignorados (FISSETTE, 2010). A Figura 3.3 apresenta a margem máxima e os vetores de suporte.

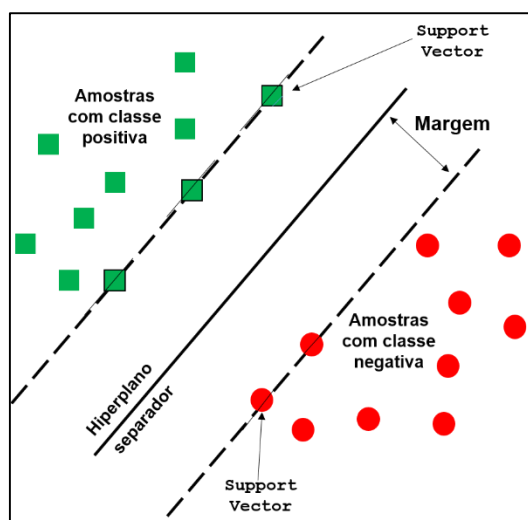


Figura 3.3 – Hiperplano separador e Vetores de Suporte

Os casos de SVMs lineares com margens rígidas que definem fronteiras lineares a partir de dados linearmente separáveis, e estão baseados nos seguintes conceitos matemáticos: Seja  $T$  um conjunto de treinamento com  $n$  dados  $x_i \in X$  e seus

respectivos rótulos  $y_i \in Y$ , em que  $X$  constitui o espaço dos dados e  $Y = \{-1, +1\}$ .  $T$  é linearmente separável se é possível separar os dados das classes  $+1$  e  $-1$  por um hiperplano (LORENA; DE CARVALHO, 2007). A equação desse hiperplano é apresentada na Equação 1, em que  $w \cdot x$  é o produto escalar entre os vetores  $w$  e  $x$ ,  $w \in X$  é o vetor normal ao hiperplano descrito e  $b/\|w\|$  corresponde à distância do hiperplano em relação à origem, com  $b \in \mathcal{R}$

$$H(w, b) = w \cdot x + b = 0 \quad (1)$$

Por outro lado a distância euclidiana de um ponto até o hiperplano é dada pela equação (2):

$$d(x, H(w, b)) = \frac{w^T * x + b}{\|w\|} \quad (2)$$

Verifica-se que a maximização da margem do hiperplano de separação pode ser obtida pela minimização de  $\|w\|$ . Estamos diante de um problema de otimização quadrático, cuja solução possui uma ampla e estabelecida teoria matemática que pode ser revisada em (LORENA; DE CARVALHO, 2007; GUGGENBERGER, 2012; HSU; CHANG; LIN, 2016).

O modelo original dos SVMs foi desenvolvido para classificação binária ( $k=2$ ) com um critério de maximização da margem embora os problemas reais com frequência precisem ser classificados em mais de duas classes (FISSETTE, 2010). De fato, o problema de classificação multi-classe ( $k>2$ ) são comumente descompostos em vários subproblemas binários, permitindo que os SVM tradicionais possam ser aplicados diretamente. Duas abordagens para descompor em problemas binários são o *one-versus-rest* (1VR) e *one-versus-one* (1V1) (WANG; XUE, 2014) .

Em (CRAMMER; SINGER, 2002) propõe-se uma nova abordagem direta para resolver o problema multiclasse sem precisar da decomposição em classificadores binários mas de um processamento único de otimização. Contudo, de acordo com a comparação feita em (HSU; LIN, 2002), os métodos de decomposição são mais adequados para usos práticos.

Existem diferentes implementações de SVM, entre as principais estão SMO (*Sequential Minimal Optimization*), LIBSVM e LIBLINEAR. O algoritmo SMO divide o problema de programação quadrática em várias problemas menores (PLATT, 1999).

O LibSVM é o principal software SVM utilizado na atualidade (CHANG; LIN, 2013) e apoia às tarefas de classificação binária e multiclasse como, por exemplo, os métodos de regressão. O LIBLINEAR é uma biblioteca *open-source* para classificação linear em grande escala e está inspirada altamente pelo LibSVM. Além disso, experimentos demonstram que é mais rápido que os classificadores lineares já existentes (FAN et al., 2015), embora sua maior utilidade esteja na classificação de documentos (*document data*). Por exemplo ele gasta só alguns segundos na fase de treinamento para a classificação do texto do *Reuters Corpus Volume 1* que possui mais de 600,000 documentos.

No projeto foram avaliados os resultados dos pacotes LibSVM e LIBLINEAR aplicados aos *tweets* de treinamento e teste.

### 3.1.3 Natural Language Processing (NLP)

O Processamento de Linguagem Natural (NLP) é uma ampla área de pesquisa em Inteligência Artificial que estuda como os computadores podem ser utilizados para analisar e manipular as linguagens naturais dos humanos que se apresentam nos textos (VIJAYARANI; ILAMATHI; NITHYA, 2015) que podem ser orais ou escritos (LIDDY, 2001). As principais aplicações práticas do NLP são o Reconhecimento de fala (*Speech processing*), extração de relações (*Relation extraction*), classificação de documentos (*Document Categorization*), simplificação de texto (*summarization of text*) e análise de sentimento (*Sentiment Analysis*) (REESE, 2015).

As tecnologias utilizadas para a análise de fala e linguagem se apoiam em modelos ou representações formais do conhecimento aos níveis da linguagem: fonologia, fonética, morfologia, sintaxe, semântica, pragmática e discurso. Estes modelos podem ser máquinas de estado, sistemas de regras, modelos lógicos e estatísticos (JURAFSKY; MARTIN, 2006).

A fonologia é o ramo da linguística que estuda o sistema sonoro de um língua e preocupa-se com a maneira como os sons se organizam dentro dela, classificando-os em unidades capazes de distinguir significados, chamadas fonemas. A morfologia estuda a estrutura e formação das palavras por meio de elementos morfológicos como radical, afixo, vogal ou consoantes de ligação. A sintaxe é o conhecimento das estruturas das relações entre as palavras para formar orações ou frases. A semântica

determina os possíveis significados de uma oração de acordo as interações entre as palavras, esse nível inclui a desambiguação de palavras que possuem mais de um significado (LIDDY, 2001; JURAFSKY; MARTIN, 2006). A pragmática preocupa-se com o uso da linguagem para o falante expressar suas intenções e propósitos em determinadas situações, portanto o contexto é necessário para o entendimento do texto (LIDDY, 2001). Finalmente, o nível de discurso está focado no entendimento de textos maiores do que uma simples oração (LIDDY, 2001; JURAFSKY; MARTIN, 2006).

### 3.1.4 Pré-processamento

O pré-processamento é a fase em que os dados são preparados para a mineração ou para serem utilizados em qualquer projeto de *Machine Learning* (VIJAYARANI; ILAMATHI; NITHYA, 2015). As técnicas de pré-processamento de dados quando são aplicadas podem melhorar consideravelmente a qualidade total dos padrões descobertos e/ou do tempo necessário para a mineração. Existem muitos fatores que constituem a qualidade dos dados como a precisão, integralidade, consistência, prontidão, credibilidade e interpretabilidade. A falta de precisão dos dados pode ter muitas razões, por exemplo que os usuários expressamente submetem informação inexata porque não desejam submeter informação pessoal (como data de nascimento); o caso da falta de integralidade e consistência dos dados acontece, por exemplo, quando dados mandatórios simplesmente não foram incluídos ou foram submetidos em um formato errado (como erro em campos de data por causa do formato); a prontidão dos dados é importante porque supor que um usuário está avaliando as vendas do final do mês mas alguns dos vendedores não enviaram seus registros de vendas, isto causaria muitos ajustes e correções posteriores; a credibilidade reflete quanta confiança tem os usuários nos dados; e a interpretabilidade reflete se os dados são fáceis de entenderem ou não (J.HAN, J.PEI, 2012).

Os principais tipos de processamento de dados são limpeza, integração, transformação e seleção. A seguir são explicados os três tipos utilizados no presente projeto:

Limpeza de dados ou *Data Cleaning*: Os dados que se encontram no mundo real tendem a ser incompletos, ruidosos e inconsistentes. As rotinas de limpeza de dados tentam preencher valores faltantes, suavizar os dados eliminando ruído e detectando *outliers* (variação ou erro aleatório observado em uma variável medida) e corrigir inconsistências presentes nos dados. Se os usuários acham que os dados estão “sujos”, é pouco provável que eles confiem nos resultados de qualquer algoritmo de mineração aplicado (J.HAN, J.PEI, 2012).

Transformação de dados ou *Data Transformation*: Consiste em transformar ou consolidar os dados em um formato mais adequado para mineração porque permite entender com maior facilidade os padrões descobertos. Vários tipos de transformação são possíveis, mas para o projeto foi utilizado a técnica de Normalização que consiste em escalonar os atributos ou *features* para um novo intervalo mais adequado para o algoritmo de mineração a ser aplicado. No caso das unidades de medida usadas, elas podem afetar a análise dos dados, por exemplo alterando as unidades de metros para polegadas para o caso da altura, ou de quilos para libras para o caso do peso, podem conduzir a resultados muito diferentes. Para evitar a dependência na escolha de uma unidade de medida, os dados devem estar normalizados ou padronizados (J.HAN, J.PEI, 2012) porque isso faz que todos os *features* possuam o mesmo ‘peso’. Adicionalmente, para as tarefas de classificação, quando os *features* têm valores não-padronizados muito grandes, o treinamento do algoritmo pode levar muito tempo (FAN et al., 2015). Dessa maneira, recomenda-se transformar os dados a um intervalo  $[-1, 1]$  ou  $[0.0, 1.0]$ .

Por outro lado, o pré-processamento focado na mineração de textos (*text mining*) permite reduzir o tamanho dos documentos por meio de técnicas como a remoção de *stop words*, *stemming* e *tokenization* (VIJAYARANI; JANANI, 2016).

O *stemming* é um método utilizado para identificar a raiz ou *stem* de uma palavra. Por exemplo, as palavras em inglês “connect”, “connected”, “connecting”, “connections”, todas elas podem ser reduzidas à palavra “connect”, (RAMASUBRAMANIAN; RAMYA, 2013). O objetivo do stemming é remover vários sufixos e assim reduzir o número de palavras e poupar memória do computador e tempo de processamento. Um dos algoritmos mais utilizados para esta atividade é o algoritmo de Porter (VIJAYARANI; JANANI, 2016).

A remoção de *stop words* consiste em remover as palavras comuns que têm pouca probabilidade de ajudar na mineração de textos e para reduzir a dimensionalidade do espaço de *features*. Alguns exemplos são as preposições, os artigos e os pronomes. A decisão das palavras a serem eliminadas depende do domínio do problema. Exemplos de *stop words* em inglês: “a”, “is”, “you”, “an” (VERMA, 2014).

O processo de *tokenization* consiste em quebrar o texto em frases, palavras, símbolos ou elementos com sentido chamados “*tokens*” com o objetivo de analisar as palavras de uma oração e identificar quais são consideradas como “chaves” (*keywords*). Esse processo poderia ser considerado como trivial porque aparentemente consiste só em armazenar o texto em um formato que possa ser entendido pelo computador. Contudo, devem ser aplicadas algumas regras de acordo com a necessidade do problema, por exemplo a remoção ou não de sinais de pontuação como parêntese, hífen, apóstrofo entre outros; ou a transformação das abreviações e acrônimos em sua forma padrão (VERMA, 2014). Na Figura 3.4, apresenta-se um exemplo prático de *tokenization* com o espaço em branco como separador de *tokens*.

Comumente, o processo de *tokenization* pode ser usado em textos escritos em por exemplo inglês, espanhol, português ou francês porque essas línguas usam o espaço em branco para separar as palavras, mas não em línguas como o chinês, hindi, tailandês. Portanto, ainda é uma necessidades desenvolver ferramentas de *tokenization* para as demais línguas (VIJAYARANI; JANANI, 2016)

Língua: Inglês
Caráter separador: espaço em branco
Texto Entrada: I am tired of being BULLIED and RIDICULED
Texto Saída : I, am, tired, of, being, bullied, and, ridiculed

Figura 3.4 – Exemplo do processo de *Tokenization*

Finalmente, apresenta-se a seleção de termos ou *feature selection* que pode ser realizado por por medidas estatísticas simples, como a frequência de termo, conhecida como TF (do inglês *term frequency*), e frequência de documentos, conhecida como DF (do inglês *document frequency*). A frequência de termo

contabiliza a frequência absoluta de um determinado termo ao longo da coleção textual. A frequência de documentos, por sua vez, contabiliza o número de documentos em que um determinado termo aparece. Destaca-se o critério TF-IDF (*Term Frequency Inverse Document Frequency*) como critério de ponderação e normalização leva em consideração tanto o valor de TF quanto o valor de DF (REZENDE; MARCACINI; MOURA, 2011). O TF-IDF é uma medida para avaliar a importância de uma palavra de um documento ou texto dentro de uma coleção de documentos (DINAKAR et al., 2012).

### 3.1.5 Representação de dados

A escolha da forma de representar os dados é um passo importante antes que qualquer modelagem seja realizada e deve ser muito específica ao domínio do problema. Os dados podem ser representados por um número fixo de atributos ou *features* que podem ser binários, categóricos ou contínuos. Um *feature* é sinónimo de variável de entrada (GUYON; ELISSEFF, 2006). Um exemplo da representação de dados é a utilizada na famosa base de dados de plantas Iris<sup>1</sup>, que possui 150 instâncias (50 exemplos de flores) e cada uma é representada por 4 *features* com seus valores em centímetros (*tamanho e largura da sépala, tamanho e largura da pétala*) e associada à classe que pertence. Na tabela 3.1, apresentam-se algumas instâncias da base de dados Iris.

Tabela 3.1 – Exemplos de instâncias de plantas da base de dados Iris

Instância	Tamanho sépala	Largura sépala	Tamanho pétala	Largura pétala	Classe
#005	5.3	3.7	1.5	0.2	Iris-setosa
#008	5.0	3.3	1.4	0,2	lis-setosa
#070	7.0	3.2	4.7	1.4	Iris-versicolor
#145	5.8	2.7	5.1	1.9	Iris-virginica

Na etapa de pré-processamento se encontra a principal diferença entre os processos de Mineração de Textos e outros processos de mineração de dados: a estruturação dos textos em um formato adequado para a extração de conhecimento. Para o projeto, os textos, que seriam os *tweets*, estão escritos em língua natural,

<sup>1</sup> Iris Data Set - <http://archive.ics.uci.edu/ml/datasets/Iris>

inerentemente não-estruturados, e precisa-se extrair uma representação estruturada, concisa e manipulável por algoritmos de classificação. Essas *features* poderiam ser uma única palavra ou uma sequência de palavras. Aparece o conceito de *n-gram* que são termos formados por mais de um elemento, porém com um único significado semântico (REZENDE; MARCACINI; MOURA, 2011). Os seguintes exemplos na língua inglesa representam os N-grams, um *1-gram* ou unigram é uma palavra isolada como “*turn*”, um *2-gram* ou *bigram* consiste em duas palavras sucessivas como “*turn your*”, or “*your homework*” e um *3-gram* ou *trigram* consiste em 3 palavras sucessivas como “*please turn your*” ou “*turn your homework* (JURAFSKY; MARTIN, 2015).

### 3.1.6 Medidas úteis do software WEKA

O software de mineração de dados utilizado no projeto para a criação dos modelos de classificação foi o WEKA<sup>2</sup> versão 3.8.1 (Figura 3.5).

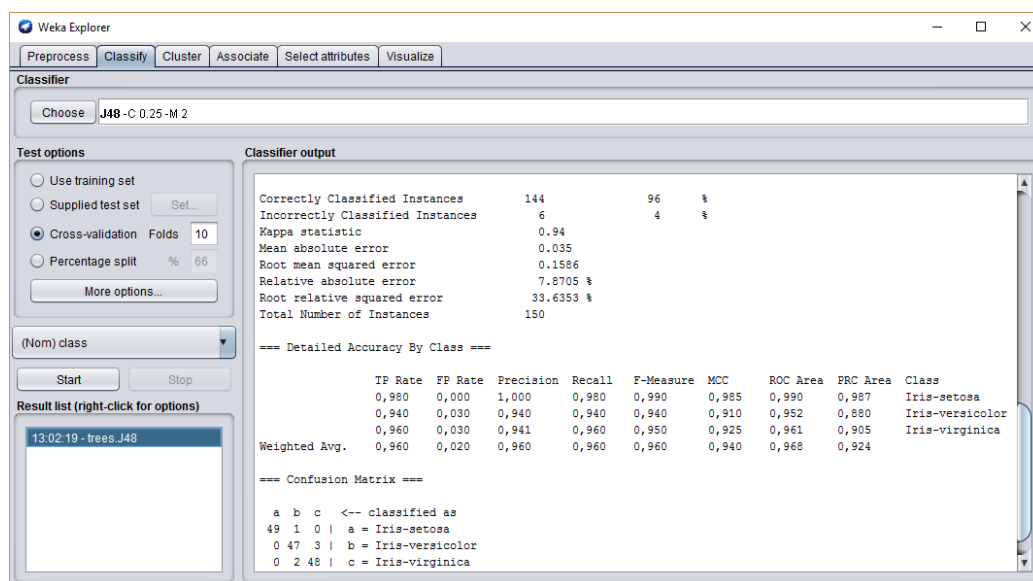


Figura 3.5– Entorno de Classificação do software WEKA 3.2.1

Após o pré-processamento, define-se a forma de teste aplicado ao *training set*. Para o projeto utilizou-se o *10-fold cross-validation* que consiste em dividir os dados aleatoriamente em 10 partes iguais. A classificação é executada 10 vezes, cada vez se utiliza uma parte diferente como *test data* enquanto as 9 partes restantes são utilizadas como *training data*. Deste modo, cada parte é usada uma *test data* uma

<sup>2</sup> Weka 3 - <http://www.cs.waikato.ac.nz/ml/weka/>



única vez. Os resultados das 10 execuções de classificação são avaliados para obter o valor médio. (FISSETTE, 2010). A seguir se apresentam as medidas utilizadas no WEKA para avaliar os resultados dos algoritmos de classificação.

- **Instâncias classificadas corretamente:** É o número e a porcentagem de instâncias/amostras classificadas corretamente a sua classe rotulada.
- **Instâncias classificadas incorretamente:** É o número e a porcentagem de instancias/amostras que foram classificadas incorretamente, se obtém da subtração de 100 – a “porcentagem das instancias classificadas corretamente”
- **Coeficiente *kappa*:** O *kappa* é uma medida de concordância entre as predições de um classificador com a classe correta. O *kappa* tem como valor máximo o 1, onde este valor 1 representa total concordância e os valores próximos e até abaixo de 0, indicam nenhuma concordância, ou a concordância foi exatamente a esperada pelo acaso. Em (LANDIS; KOCH, 1977) sugerem a interpretação do *kappa* apresentada na Tabela 3,2. Essa avaliação de concordância por meio do *kappa* é utilizada quando as escalas são categóricas e sempre quando estamos comparando dois ou mais juízes/testes. Quando um classificador tem uma alta porcentagem de instancias classificadas corretamente, o *kappa* terá um valor maior (FISSETTE, 2010).

Tabela 3.2 – Níveis de concordância utilizando o *kappa* (LANDIS; KOCH, 1977)

Valor do <i>kappa</i>	Nível de concordância
0,00 <	Pobre
0,01 – 0,20	Leve
0,21 – 0,40	Aceitável
0,41 – 0,60	Moderada
0,61 – 0,80	Considerável
0,81 – 1,00	Quase perfeita

- **Erro médio absoluto (MAE) e Raiz do erro médio quadrático (RMSE):** O MAE mede quanto próximo (distancia em valor absoluto) está a predição do classificador à classe correta e o RMSE é um forma diferente para calcular o MAE mas usando a diferença dos quadrados (FISSETTE, 2010).

- **Erro relativo e Raiz do erro relativo quadrático:** Medidas expressadas em percentagem. Menores percentagens indicam um melhor rendimento dos algoritmos (FISSETTE, 2010).
- **Matriz de confusão (*Confusion Matrix*):** É uma ferramenta útil para indicar o desempenho de um classificador para amostras de diferentes classes. Se o problema possui  $m$  classes, a matriz de confusão será  $m \times m$ . Na Tabela 3.6 apresenta-se um modelo de matriz de confusão com só duas classes “*Sim*” e “*Não*” (J.HAN, J.PEI, 2012). Uma entrada na célula  $(i, j)$  indica o número de amostras da classe  $i$  que foram classificadas como classe  $j$ . A situação ideal é que todas as amostras fora diagonal principal da matriz tenham valor ou próximo a zero. Entre os resultados do *WEKA* temos o “*TP Rate*”, “*FP Rate*”, “*Precisão*”, “*Recall*” e “*F-measure*” que são explicados por meio da matriz.

		Classe predita ( $j$ )		
		<i>Sim</i>	<i>Não</i>	<i>Total</i>
Classe correta ( $i$ )	<i>Sim</i>	TP	FN	P
	<i>Não</i>	FP	TN	N
	<i>Total</i>	P'	N'	P+N

		Classe predita ( $j$ )		
		<i>Sim</i>	<i>Não</i>	<i>Total</i>
Classe correta ( $i$ )	<i>Sim</i>	6954	46	7,000
	<i>Não</i>	412	2588	3,000
	<i>Total</i>	7366	2634	10,000

*Diagonal principal*

Figura 3.6– Matriz de confusão para as classes “*Sim*” e “*Não*”

**Positivos Verdadeiros (TP Rate):** Número de amostras da classe “*Sim*” que foram corretamente classificadas.

**Negativos Verdadeiros (TN):** Número de amostras da classe “*Não*” corretamente classificadas.

**Falsos Positivos (FP Rate):** Número de amostras da classe “*Não*” que foram classificadas classe “*Sim*”.

**Falsos Negativos (FN Rate):** Número de amostras da classe “*Sim*” que foram classificadas classe “*Não*”.

**Precisão:** É uma medida de exatidão ou *exactness*. Se obtém por meio do cálculo de  $TP/(TP+FP)$ . Corresponde à percentagem de amostras rotuladas como “*Sim*” que são efetivamente da classe “*Sim*”.

**Sensitividade ou Recall:** É uma medida de integralidade ou *completeness*. Se obtém por meio do cálculo de  $TP/(TP+FP)$ . Corresponde à porcentagem de amostras “Sim” que são corretamente classificadas.

**F-measure:** É uma alternativa para usar a precisão e a sensitividade em uma medida única. Se os valores para precisão e a sensitividade é alto, o *F-measure* será também alto (J.HAN, J.PEI, 2012) (FISSETTE, 2010) .

### 3.1.7 Teoria sobre Lógica *Fuzzy* ou *Fuzzy Logic*

A Lógica *Fuzzy* (Lógica Difusa ou Lógica Nebulosa) é um método de *Soft Computing* que explora a tolerância às imprecisões, incertezas e verdades parciais do mundo real para atingir robustez e flexibilidade. Em um sentido mais específico, Lógica *Fuzzy* é uma extensão da lógica multivalor cujo objetivo é obter um raciocínio aproximado mais do que uma solução exata. Ao contrário da lógica tradicional *Crisp*, que trabalha com valores binários (verdadeiro é 1 e falso é 0), as variáveis em Lógica *Fuzzy* podem ter um valor verdadeiro que varia dentro de um grau 0 (completamente falso) e 1 (completamente verdadeiro) em vez de descrever um “*Sim*” ou “*Não*” absoluto, isto explica-se por meio de uma função de pertinência onde qualquer valor no intervalo indica o grau de verdade (SIVANANDAM; SUMATHI; DEEPA, 2007) (MELO; CAVALCANTI, 2012). A seguir, se apresentam alguns dos conceitos envolvidos na Lógica *Fuzzy*:

- **Conjunto *Fuzzy* ou *Fuzzy Set***

O conjunto *Fuzzy* é uma extensão do clássico conjunto *Crisp* que possui uma função de pertinência com lógica binária: elemento pertence ou não pertence a esse conjunto. O conjunto *Fuzzy* fornece meios para modelar a incerteza associada com ambiguidade, imprecisão e falta de informação sobre um problema que frequentemente se encontra na vida real (WANG, 2015). Por exemplo, considera-se o significado para o conceito de “persona baixa”, para uma indivíduo “X”, esse conceito poderia ser associado a alguém com uma altura menor à 1.30m, para outro indivíduo “Y” poderia significar alguém com uma altura menor ou igual do que 1.20m. A mesma palavra “baixa” pode ter diferentes significados dependendo da percepção de cada

indivíduo, então não existe uma definição única para esse termo, além disso é um termo chamado de “variável linguística” que representa a imprecisão de nossa linguagem natural ao falar (SIVANANDAM; SUMATHI; DEEPA, 2007). O conjunto *Fuzzy* permite que os membros pertençam a um grupo em um grau de pertinência que varia no intervalo  $[0,1]$ . Na Figura 3.7, apresenta-se a variável temperatura em graus  $F^\circ$  dividida em três categorias: *Low* ( $0F^\circ - 30F^\circ$ ), *Medium* ( $30F^\circ - 70F^\circ$ ), *High* ( $70F^\circ - 120F^\circ$ ) desde o ponto de vista do conjunto *Crisp* e o conjunto *Fuzzy*. No conjunto clássico da Figura 3.7a, as três categorias tem seus limites bem definidos, portanto a temperatura pode ser classificada unicamente em uma delas, por enquanto, na Figura 3.8b os limites das categorias são nebulosos (BAI; WANG, 2007).

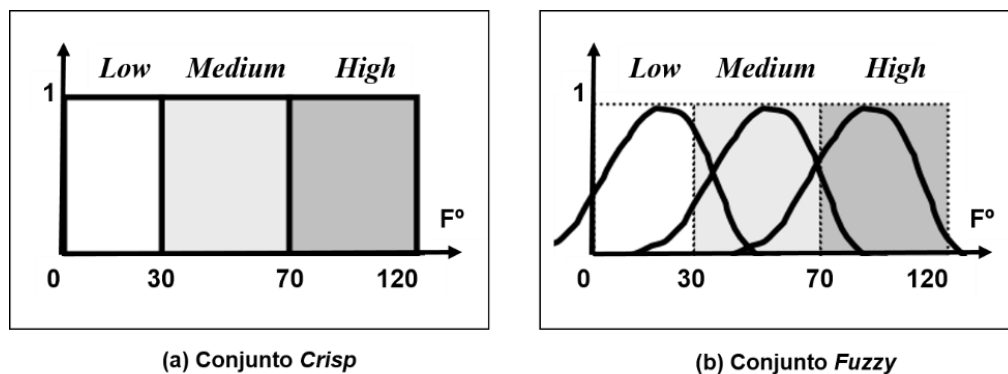


Figura 3.7– Conjunto *Crisp* vs Conjunto *Fuzzy*

Determinar a função de pertença é subjetivo porque depende da percepção que tem os indivíduos sobre os dados do problema e não de questões aleatórias. Pode-se definir então que um elemento do universo, “ $x$ ” é um membro do conjunto *Fuzzy* “ $A$ ”, e o valor de pertinência é dado pelo função  $\mu_A(x) \in [0, 1]$ . (SIVANANDAM; SUMATHI; DEEPA, 2007).

- **Função de pertença ou *Membership Function***

A função de pertença é uma curva que define a característica do conjunto *Fuzzy* por meio da atribuição de um valor o grau de pertinência para cada elemento do conjunto que varia de  $[0,1]$ . A Figura 3.8 apresenta uma curva genérica de uma função de pertinência onde o eixo horizontal representa a variável de entrada “ $x$ ” e o eixo vertical define o valor de pertinência  $\mu(x)$  dessa variável “ $x$ ”. As representações gráficas das funções de pertinência podem ser de diferentes formas como triângulos,

trapezoides, Gaussiana, entre outras e podem ser simétricas ou assimétricas (WANG, 2015).

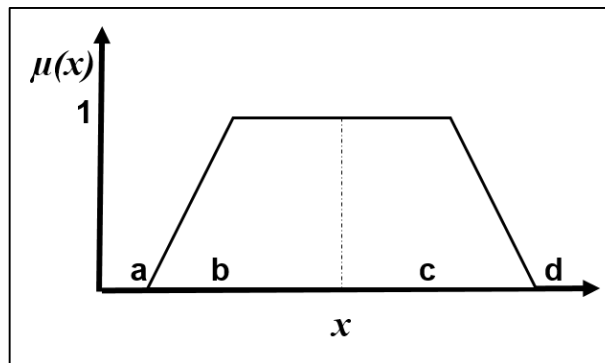


Figura 3.8– Exemplo simples de função de pertinência

- **Sistema baseado em regras Fuzzy ou *Fuzzy Rule-Based System***

As regras formam a base da lógica *Fuzzy* para obter a saída *Fuzzy* (*Fuzzy output*), elas constituem o sistema baseados em regras que é considerado como o conhecimento de um experto sobre um determinado campo ou domínio de aplicação. A lei para desenvolver um conjunto de regras *Fuzzy* está baseado no conhecimento humano (BAI; WANG, 2007). A forma das regras utiliza variáveis linguísticas como seus antecedentes e consequentes. Os antecedentes expressam uma inferência ou desigualdade a qual deve ser satisfeita. Os consequentes são o resultado da inferência. Para o caso do sistema baseado em regras *Fuzzy*, eles usa mas regras na forma *IF-THEN* dado por *IF* antecedente, *THEN* consequente (SIVANANDAM; SUMATHI; DEEPA, 2007). Por exemplo, continuando com a variável de entrada “temperatura”, poderíamos ter a seguinte regra *IF-THEN*:

*IF* a ‘temperatura’ é *LOW*, *THEN* o ‘aquecedor do motor’ deveria girar *FAST*

O resto de regras devem seguir a mesma estratégia que é muito similar à intuição humana (BAI; WANG, 2007). De fato, o conjunto de regras devem ter as seguintes propriedades (SIVANANDAM; SUMATHI; DEEPA, 2007):

- Integralidade ou *Completeness*: Um conjunto de regras *IF-THEN* é completo se qualquer combinação das variáveis de entrada possui uma apropriada variável de saída.

- Consistência ou *Consistency*: Um conjunto de regras *IF-THEN* é consistente se duas regras com os mesmos antecedentes geram diferentes resultados em os consequentes.

- Continuidade ou *Continuity*: Um conjunto de regras *IF-THEN* é contínuo se não tem regras vizinhas com conjuntos *Fuzzy* de saída que possuam intersecção vazia.

- **Sistema de inferência *Fuzzy* ou *Fuzzy Inference System***

O sistema de inferência *Fuzzy* (FIS) consiste em cinco blocos funcionais como se apresenta na Figura 3.9: uma interface de Fuzzificação ou *Fuzzification* que transforma as entradas *Crisp* em valores *Fuzzy*; um conjunto de regras que contém um número de regras *Fuzzy IF-THEN*; uma base de dados que define as funções de pertinência dos conjuntos *Fuzzy* usados nas regras; uma unidade de tomada de decisões que executa as operações de inferência nas regras; e uma interface de Defuzzificação ou *Defuzzification* que transforma os resultados *Fuzzy* em uma saída *Crisp*. (SIVANANDAM; SUMATHI; DEEPA, 2007).

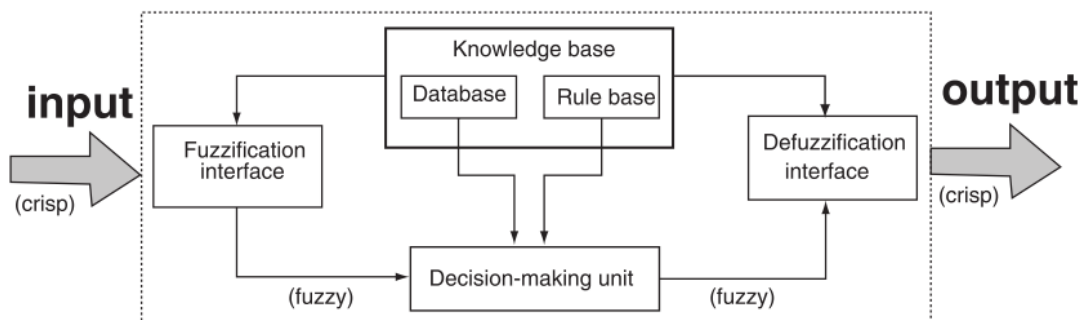


Figura 3.9– Sistema de inferência *Fuzzy* – (SIVANANDAM; SUMATHI; DEEPA, 2007)

A seguir, apresenta-se na Figura 3.10, o clássico exemplo prático para um melhor entendimento do processo da lógica *Fuzzy*: O problema da Gorjeta (*The Tipping Problem*)<sup>3</sup>. As variáveis de entrada são duas “qualidade da comida” e “qualidade do serviço”, ambas estão no intervalo [0,10] onde 10 representa a máxima qualidade; a variável de saída é o “valor da gorjeta” em porcentagem (%); as funções de pertinência para as variáveis de entrada são “Ruim”, “Decente” e “Excelente” e para a variável de saída são “Low”, “Medium” e “High”, logo elas serão variáveis linguísticas possuem um intervalo (para o caso da variável linguística “Ruim”, ela varia de [0, 4.5])

<sup>3</sup> [http://pythonhosted.org/scikit-fuzzy/auto\\_examples/plot\\_tipping\\_problem.html](http://pythonhosted.org/scikit-fuzzy/auto_examples/plot_tipping_problem.html)

(BAI; WANG, 2007). As regras *Fuzzy* definem a relação entre as variáveis de entrada e saída.

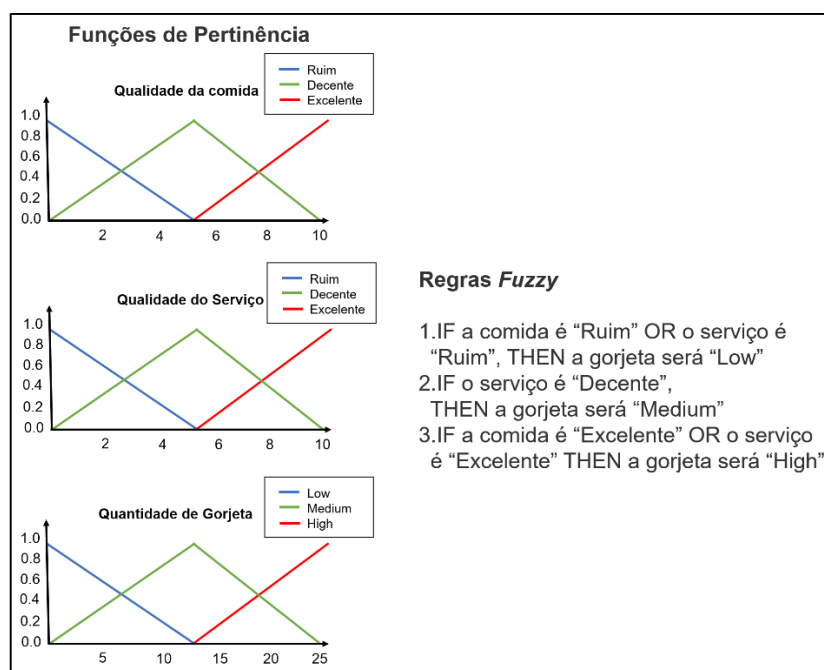


Figura 3.10– Funções de Pertinência e Regras *Fuzzy* do problema da Gorjeta

O processo começa com a *Fuzzificação* das variáveis de entrada (por exemplo, qualidade da comida= 6.5 e qualidade do serviço=9.8, portanto ambas pertencem as variáveis linguísticas de "Decente" e "Excelente") que consiste em transformá-las em valores que variam no intervalo [0,1] por meio das funções de pertinência. Logo, se aplicam as regras *Fuzzy* com as variáveis linguísticas obtidas e todas as saídas das regras ativadas (para nosso exemplo, unicamente se ativaram as regras 2 e 3) devem ser combinadas para obter a distribuição Fuzzy de saída como se mostra na Figura 3.11, esta fase é chamada de "processo de agregação". Finalmente, deve-se obter um número *Crisp* por meio de métodos de *Defuzzificação*, o mais comum é o *Center of mass*. Este método utiliza a distribuição de saída e encontra seu centro de massa que seria o número *Crisp*, que para o exemplo a gorjeta resulta ser de 20.2% (SIVANANDAM; SUMATHI; DEEPA, 2007).

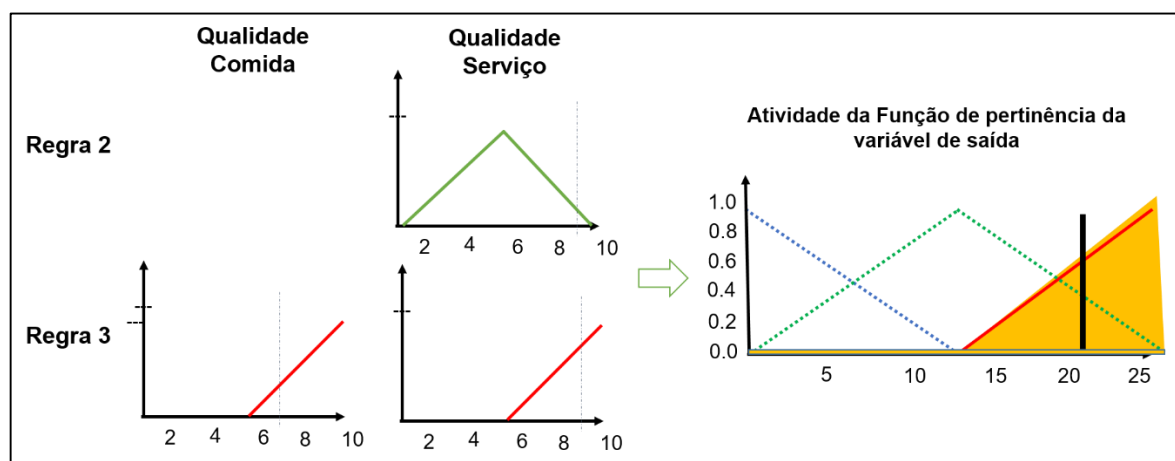


Figura 3.11– Combinando saídas das regras em uma distribuição de saída

- **Métodos de Inferência *Fuzzy* ou *Fuzzy Inference Methods***

O método de inferência *Fuzzy* mais importante e mais utilizado é o *Mamdani*. Este método foi desenvolvido pelo Mamdani e Assilian em 1975. Outro método conhecido é o *Sugeno* ou *Takagi-Sugeno-Kang*. A principal diferença entre esses métodos está nos consequentes das regras *Fuzzy*. Os sistemas *Fuzzy Mamdani* utilizam conjuntos *Fuzzy* como consequentes, enquanto os sistemas *Fuzzy Sugeno* utilizam funções lineais das variáveis de entrada como os consequentes. (SIVANANDAM; SUMATHI; DEEPA, 2007; WANG, 2015). O sistema *Fuzzy* desenvolvido no projeto utilizou o método de inferência Mamdani.

- ***Mamdani's Fuzzy Inference Method***

O *Mamdani* é o mais visto na metodologia *Fuzzy* porque é mais intuitivo. Após do processo de agregação, tem um conjunto *Fuzzy* para cada variável de saída que precisa passar pelo processo de *Defuzzificação*. Na Figura 3.12, apresenta-se um exemplo de um sistema de inferência *Fuzzy* com método *Mamdani* modelado em MATLAB<sup>4</sup>, com três variáveis de entrada e uma variável de saída chamada “severidade” (*severity*).

A seguir, apresenta-se os seis passos para calcular a saída do FIS utilizando as variáveis de entrada (SIVANANDAM; SUMATHI; DEEPA, 2007):

1. Determinar o conjunto de regras *Fuzzy*.
2. Processo de Fuzzificação.

<sup>4</sup> MATLAB – <https://es.mathworks.com/products/matlab.html>



3. Combinação das entradas *fuzzificadas* de acordo as regras *Fuzzy* para estabelecer o peso das regras.
4. Encontrar o consequente da regra por meio do peso dela e a função de pertinência da saída.
5. Combinação dos consequentes para obter a distribuição da saída.
6. Processo de Defuzzificação da distribuição de saída (este passo é unicamente se é necessário obter a classe *Crisp*).

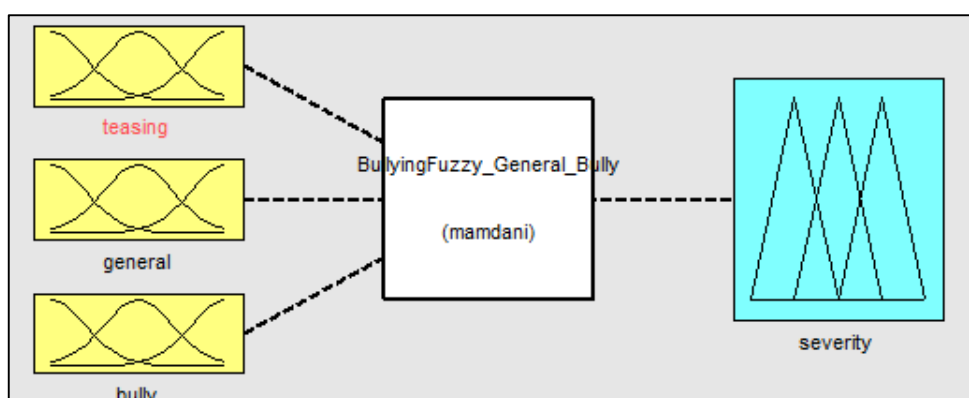


Figura 3.12– Exemplo de Sistema de Inferência *Fuzzy* com método *Mamdani*

## 3.2 Revisão de trabalhos

Em (AJAO; HONG; LIU, 2015) menciona-se que a popularidade da rede social *Twitter*, tem feito que ele possua muita influência nas comunicações do dia a dia, fortaleça as relações sócias e a disseminação da informação. Os textos publicados no *Twitter*, *tweets*, são mensagens de máximo 140 caracteres e devido ao espaço limitado, os usuários utilizam um dicionário informal de palavras que unicamente é usado nas redes sócias que inclui abreviações, erros tipográficas, uso de *emoticons*<sup>5</sup>, ironias, sarcasmos e temas de maior tendência por meio de *hashtags*<sup>6</sup>. A coleta de *tweets* está disponível para acessar de forma pública por meio do *Twitter REST API*<sup>7</sup> o que faz que o *Twitter* seja uma poderosa ferramenta para a análise e monitoramento

<sup>5</sup> *Emoticon* - <https://pt.wikipedia.org/wiki/Emoticon>

<sup>6</sup> *Hashtag* - <https://pt.wikipedia.org/wiki/Hashtag>

<sup>7</sup> *Twitter Developer Documentation* - <https://dev.twitter.com/rest/public>

das opiniões dos usuários. Nessa pesquisa, também menciona-se que existe um crescimento dos casos reportados de perseguição (*stalking*) e *cyberbullying*, em que na maioria dos casos são feitos por usuários anônimos o que dificulta a coleta de evidência suficiente para processar aos criminosos. Contudo, é importante desenvolver métodos para a detecção dos textos considerados como *cyberbullying* porque existem casos em que as vítimas se sentem muito prejudicadas e podem ser conduzidas a um eventual suicídio (HINDUJA; PATCHIN, 2010).

Sobre os trabalhos de detecção de *cyberbullying* em redes sociais, temos a pesquisa de (NALINI; SHEELA, 2015) que propõe uma abordagem para detectar mensagens de *cyberbullying* publicados em *Twitter* por meio de um esquema de pesos TF-IDF de seleção de *features* ou atributos e a identificação dos agressores e vítimas mais ativos por meio da modelagem de um gráfico de *cyberbullying*. Em (KONTOSTATHIS et al., 2013) se realizou uma análise dos termos utilizados em *cyberbullying*; a fonte de dados foi da rede social Formspring.me que permite aos usuários fazerem perguntas a outros usuários e a técnica utilizada foi um algoritmo supervisionado que além de identificar termos permitiu identificar os textos como instancias de *cyberbullying*. Em (DINAKAR; REICHART; LIEBERMAN, 2011), utilizou-se 4,500 comentários do *Youtube* como fonte de dados para o problema do *cyberbullying* o qual foi descomposto em na detecção de temas sensíveis e de natureza pessoal como aparência física, sexualidade, raça, cultura e inteligência por meio de algoritmos supervisionados de classificação; a justificativa desta decomposição foi que esses temas envolvem mensagens escritas com rudeza, ignorância e profanidade. Em (PARIME; SURI, 2014) aplicou-se a técnica de análise de sentimento (*Sentiment Analysis*) em um conjunto de textos rotulados como “ausência” ou “presença” de *cyberbullying* que foram coletados da rede social *MySpace* para identificar textos ofensivos.

Em (NANDHINI; SHEEBA, 2015), propõe-se uma arquitetura para identificar a presença de termos de *cyberbullying* e classificação de atividades de *cyberbullying* em textos coletados das rede sociais Formspring.me e *MySpace* por meio de lógica *Fuzzy* e um algoritmo genético. Essas atividades são “guerra verbal” (*flaming*), assédio, racismo e terrorismo. A principal funcionalidade foi ajustar a representação da informação antes do processo de classificação utilizando como base de conhecimento as regras *Fuzzy*.

Em (BOMMERSON, 2015) enfatiza-se a importância de acompanhar o comportamento dos usuários nas redes sociais como *Twitter* porque ajuda na prevenção de *cyberbullying*. Na pesquisa tentou-se classificar automaticamente mensagens de *cyberbullying* e avaliar os resultados de 6 algoritmos de *Machine Learning* para essa atividade: *NaiveBayes*, *IBk* (método de *lazy learning*), *JRip* (método baseado em regras), *J48* (método de árvore de decisão), *SMO* e *LibSVM* que são algoritmos SVM; utilizou-se o *software Weka* o pré-processamento e classificação dos dados, e os resultados demonstraram que o *LibSVM* conseguiu obter maior precisão de classificação.

Por outro lado, o trabalho de (MANCILLA-CACERES et al., 2012) considera que obter dados das rede sociais pode não ser muito efetiva. Portanto, desenvolveu-se um jogo *online* para observar, de uma forma não intrusiva, os papéis dos participantes em uma sala de aula (agressor, vítima ou testemunha) por meio de suas interações sociais. O jogo é em equipes e tem 3 fases:

- Os usuários decidem com quem eles gostariam e não gostariam de jogar (*Escolha de membros*)
- Todos os membros da equipe devem trabalhar juntos para responder perguntas (*Tarefa de colaboração*)
- Um dos membros da equipe tem que escolher uma resposta errada (*Tarefa de competitividade*)

Nas tarefas de Colaboração e Competitividade, os participantes usam um *chat* para sua comunicação. A informação previa sobre os participantes foi coletada por meio de questionários para medir a agressão usando algumas escalas: *Agressor*, *Briga*, *Vitimização*, *Atitude positiva diante bullying* e *Necessidade pelo Controle e Domínio*. Assim, segundo análise dos resultados, um especialista coloca uma das duas etiquetas “agressor” ou “não-agressor” aos participantes. Neste estudo, utilizou-se a análise de variância ANOVA para estudar as interações e diferenças nos comportamentos dos agressores e dos não-agressores durante as tarefas de colaboração e competitividade. O resultado foi que se encontraram interessantes padrões de interação, diferentes para os agressores, vítimas e testemunhas. Adicionalmente, a data coletada durante o jogo inclui uma quantidade grande de mensagens de textos que, no futuro, pode ser utilizada por técnicas de NLP para

aumentar a eficiência da análise do comportamento do agressor e da vítima.

Continuando com as pesquisas que utilizaram jogos para combater o problema do bullying, temos a pesquisa de (CLÁUDIO et al., 2015) que desenvolveu o jogo “*StopBully*” baseado no conceito de *Serious Games*, que são softwares interativos que usam videogames com propósitos além da diversão. O objetivo do jogo é ajudar as vítimas e testemunhas a cambiarem seus comportamentos diante do *bullying* por meio de episódios simulados no jogo. Esse jogo não inclui os fatores de colaboração e competência mas se eles fossem considerados, as interações poderiam ser representados utilizando mensagens que logo seriam analisadas com algoritmos de *Machine Learning*.

A pesquisa de (FISSETTE, 2010) ainda não está focada na detecção de bullying, é útil porque tem como objetivo determinar que tipo de informação faz possível a identificação dos autores de textos curtos por meio da classificação. Essa pesquisa considerou os *emoticons* como parte dos textos por serem muito utilizados no entorno digital como parte de um processo inconsciente. A fonte de dados foi coletada de *fok.nl* que é um site de mensagens escritos em holandês sobre diferentes temas, em total foram selecionados 25 mensagens por autor (40 autores). Os *features* utilizados foram unigrams e bigrams; e os classificadores SMV utilizados foram o SMO (SVM próprio do *software WEKA*), LibSVM e LibLINEAR, sendo esse último quem conseguiu maior precisão de classificação (15.85% de amostras corretamente classificadas, embora é considerado uma porcentagem muito baixa). Contudo, a pesquisa recomenda que as ferramentas utilizadas para a detecção de autores não deveria ser considerada como evidencia definitiva mas pode ser utilizada para fornecer, por exemplo, a lista dos 5 possíveis autores de crimines cibernéticos.

As pesquisas de (XU, 2015) (XU; ZHU; BELLMORE, 2012) mostram que os textos publicados nas redes sociais são uma fonte valiosa para o estudo do *bullying*, não só do tipo *cyberbullying*, se são analisados por meio de técnicas apropriadas de *Machine Learning* e *Natural Language Processing*. Os pesquisadores disponibilizaram os conjuntos de dados e o código dos algoritmos para classificação no seu site<sup>8</sup> para a comunidade interessada no estudo científico do *bullying*. Os textos foram coletados do *Twitter*, *tweets*, e passaram por um pré-processamento chamado de “*enrichment*”

---

<sup>8</sup> *Data and code for the study of bullying*- <http://research.cs.wisc.edu/bullying/data.html>

que consistiu em considerar unicamente os *tweets* que possuam específicas palavras chaves ou *keywords* definidas pelos expertos em matéria de *bullying*. Esta tarefa foi realizada em um programa na linguagem *Python* por meio da biblioteca *Tweepy* e o *Twitter API* e a coleta de *tweets* foi desde o ano 2011 até 2013, em total se coletaram 32,477,558 *tweets* para ser parte do conjunto de teste (*test set*) e 7,321 *tweets* foram parte do conjunto de treinamento (*training set*). A representação do *tweet* foi por meio de vetores e as *features* selecionadas foram uma combinação de *unigrams+bigrams* (1g2g). O arquivo vocabulário está composto por esses *unigrams* e *bigrams* (4,524 no total) que foram encontradas nas experiências de *bullying*. Logo, após de comparar o desempenho de diversos algoritmos para classificação de texto, a escolha final foi o algoritmo SVM.

# Capítulo 4

## PROPOSTA DE CONCLUSÃO

---

---

A proposta de este projeto consiste no desenvolvimento de um *framework* que permita ao pessoal de escola determinar o nível de severidade de um episódio de *bullying*, relatado em textos curtos escritos por estudantes na faixa etária de 11 a 14 anos. Os textos possuem as mesmas características que os *tweets* do *Twitter*, alguns exemplos são apresentadas na Figura 4.1. Note-se que esses *tweets* relatam os episódios de *bullying* desde o ponto de vista das vítimas e repórteres, as palavras utilizadas expressam seus medos e as consequências de sofrerem *bullying*. Portanto, estamos diante de um problema real, que tentamos detectar por meio do *framework* desenvolvido no projeto.



Figura 4.1– Exemplos de tweets sobre episódios de bullying

Os *tweets* utilizados para o desenvolvimento do *framework* foram coletados desde junho 2016 até junho 2017, com um total de 25,087 *tweets*. Para a criação dos novos modelos de classificação, se utilizaram 6,687 *tweets* (26.6% do total de *tweets*) como *training set*; enquanto foram avaliados 18,400 *tweets* (73,3% do total de *tweets*) para testar o *framework*. Os novos textos seriam coletados por meio de uma rede social interna web chamada de “*Class21*”<sup>9</sup>, desenvolvida para o projeto e a avaliação deles seria utilizando um sistema desenvolvido em *Java Swing*, que executa os classificadores SVM de (XU, 2015) e o sistema de Lógica *Fuzzy* para obter o nível de severidade dos textos.

Na figura 4.2 se mostram as principais tarefas realizadas no projeto. As tarefas de pré-processamento dos *tweets* e a criação e avaliação dos novos modelos SVM (que inclui também um modelo utilizando a árvore de decisão J48 do *software WEKA*) permitiram comparar com os resultados obtidos em (XU, 2015) além de permitir entender todas as fases necessárias para o processo de classificação. A funcionalidade do *framework* foi desenvolvida por meio de quatro tarefas principais: **(1)** Pré-processamento dos *tweets* de teste e dos *textos* coletados por meio do site “*Class21*”, **(2)** Execução do primeiro classificador SVM “*Bullying Trace*” de (XU, 2015) que é o primeiro filtro dos textos, só aqueles que foram classificados como “yes” (significa que o texto trata-se de um episódio de *bullying*) continuam com a avaliação no *framework*. **(3)** Execução dos três classificadores SVM “*Bullying Form*”, “*Teasing Trace*”, “*Author Role*” de (XU, 2015), se as saídas deles para o texto avaliado são entradas válidas (que serão mencionada nas próximas seções) se continua com a seguinte tarefa. **(4)** Desenvolvimento do sistema de Lógica *Fuzzy* que inclui a definição das regras que permitem determinar a severidade do texto.

---

<sup>9</sup> Class21 - <http://lionlineproyectos.com/class21>

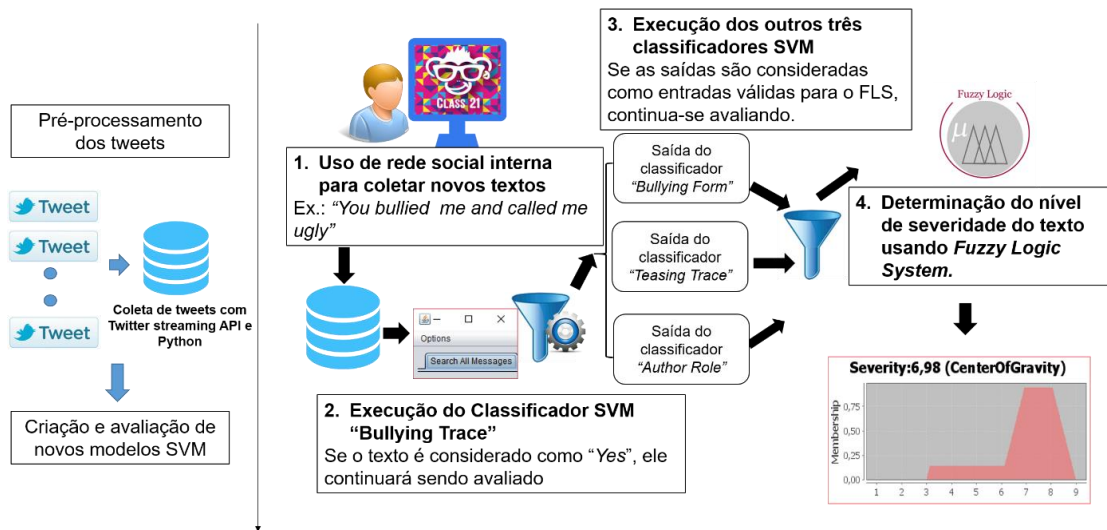


Figura 4.2– Tarefas para o desenvolvimento do *framework* para determinar o nível de severidade em textos em inglês

## 5.1 Pré-Processamento dos dados

O pré-processamento dos dados é uma fase importante para a classificação de dados, como foi mencionado no Capítulo 3. Para o projeto, a coleta de *tweets* se realizou em um programa na linguagem *Python* por meio da biblioteca *Tweepy* e o *Twitter API*, se definiu as *keywords* que deveriam possuir os textos a serem coletados. Em (XU, 2015) se colocaram algumas *keywords* e no projeto se adicionou outras mais obtidas da literatura, o conjunto total de *keywords* se mostram na Figura 4.3. As indicações foram que o *tweet* deve ter no mínimo uma palavra que comece com "bull", e não deve ser um "*re-tweet*"<sup>10</sup> ou seja não deve conter a palavra "RT".

school, victim, crying, ignored, pushed, rumors, locker, spread, shoved, rumor, teased, kicked, 'beat up', excluded, isolated, torment, terrify, violent, violence, offensive, insult, frightened, destructive, manipulate, hostile, harassed, hurt, hit, ridiculed, embarras, humiliate, degrate, called names, threat, slap, leave, steal, taunt, punch, spit, bullied, bully, bullying, bullyer, bulling, bullyied, cyberbullying, victim

<sup>10</sup> Definição de *re-tweet* - <https://www.thebalance.com/retweet-definition-what-retweet-means-and-how-to-use-them-896699>



Figura 4.3– Lista de *keywords* associadas ao contexto de *bullying* de acordo com a literatura

### 5.1.1 Tokenization

No total foram coletados 25,087 *tweets* os quais passaram por um processo de *tokenization* definido em um programa Java baseado nas regras de (XU, 2015) e melhoradas para o projeto, alguns exemplos delas são:

- Se o *tweet* tiver alguma URL (“http://” ou “https://”), ela será substituída pelo *token* “HTTPLINK”.
- Se o *tweet* tiver algum USUÁRIO (Exemplo @pamela123), ele será substituído pelo *token* “USER”.
- Se o *tweet* tem palavras de negação, elas serão substituídas pelas seguintes regras.
- Regras para as abreviaturas do inglês no tempo futuro.

### 5.1.2 Representação dos *tweets* em formato vetor

O *tweet* foi representado através de um vetor que terá a estrutura definida na Figura 4,4 (b). Os *features* (ou índices do vetor) foram *unigrams* e *bigrams* que dão um total de 4,524 elementos no arquivo vocabulário. Um *tweet* só possui um conjunto reduzido dos *features*, portanto, o vetor terá células associadas com um valor igual a zero que não serão tomadas em conta durante sua avaliação.

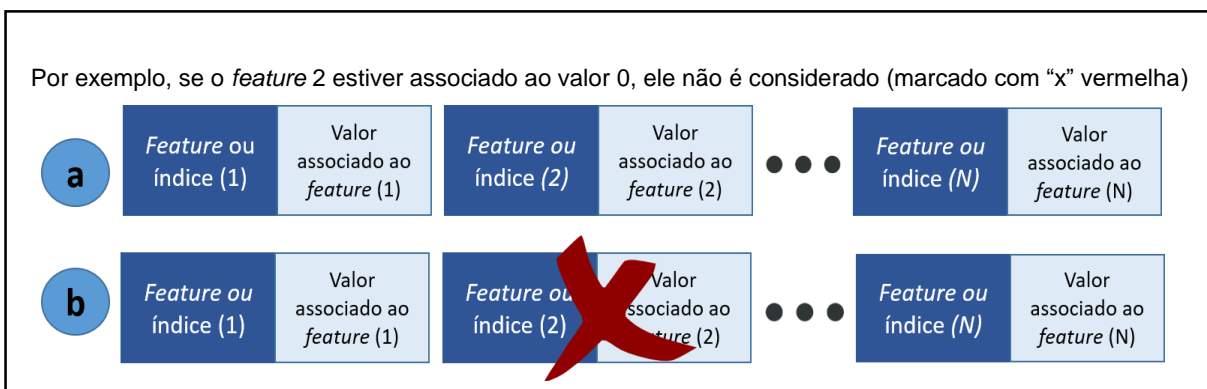


Figura 4.4– (a) Representação de texto em vetor com todos os *features*. (b) Representação de texto só com os *features* utilizados

Por exemplo, temos o texto: “*She was bullied bullied at school and after school*”. Ele possui 7 *unigrams* únicos (*she, was, bullied, at, school, and, after*) e 5 *bigrams*

(*she was, was bullied, bullied at, at school, school and*), em total haveria 12 *features* ou índices (o valor do índice é obtido do número da linha que ocupa no arquivo de vocabulário) que é apresentado na Figura 4.5. Portanto, esses índices serão os *features* úteis porque vão ter valor diferente a zero, os demais não serão de interesse.

Tweet:	
“ <i>She was bullied bullied at school and after school</i> ”	
Qual é o <i>feature</i> (número de linha) para o <b>Unigram</b> segundo o vocabulário?	Qual é o <i>feature</i> (número de linha) para o <b>Bigram</b> segundo o vocabulário?
2928 she	2933 she was
3735 was	3737 was bullied
436 bullied	444 bullied at
265 at	276 at school
2850 school	2851 school and
152 and	
81 after	

Figura 4.5 *Features* ou índices do exemplo

Continua-se com o processo do cálculo do valor do *feature* definido no passo anterior. Utilizou-se a frequência de aparição (número de vezes) do *unigram* e *bigram* dentro do texto, com o formato:

feature1:freqüência1 feature2:freqüência2 ...

Para o exemplo, o vetor seria (os detalhes estão na Tabela 4.1):

2928:1.0 81:1.0 2850:2.0 2851:1.0 436:2.0 276:1.0 2933:1.0 3735:1.0 152:1.0  
265:1.0 3737:1.0 444:1.0

Tabela 4.1 – Frequência dos *features* ou índices do exemplo

Unigram		Bigram	
<i>Feature</i>	Frequência do <i>feature</i> no tweet	<i>Feature</i>	Frequência do <i>feature</i> no tweet
2928 she	1	2933 she was	1
3735 was	1	3737 was bullied	1
436 Bullied	2	444 bullied at	1
265 at	1	276 at school	1
2580 school	2	2851 school and	1
152 and	1		
81 after	1		

Contudo, o vetor não está normalizado e é importante que os valores dos *features* estejam em um intervalo [0,1]. O vetor será normalizado com a NORMA

EUCLIDIANA que é definida com a equação (1) e em conceito significa a longitude do vector.

$$ss = \|x\|_2 = \left( \sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} \quad (1)$$

Continuando com o exemplo o  $ss$  é

$ss = \text{Raiz}(1+1+2^2+1+2^2+1+1+1+1+1+1) \rightarrow 4,24264$

O formato normalizado é:

$\text{feature1}:(\text{frequência1}/ss) \quad \text{feature2}:(\text{frequência2}/ss) \quad \dots$

Finalmente, o vetor do exemplo é apresentado na Figura 4.6

2928:0.23570226039551587	81:0.23570226039551587	2850:0.47140452079103173
2851:0.23570226039551587	436:0.47140452079103173	276:0.23570226039551587
2933:0.23570226039551587	3735:0.23570226039551587	152:0.23570226039551587
265:0.23570226039551587	3737:0.23570226039551587	444:0.23570226039551587

Figura 4.6– Vetor normalizado para o exemplo

Os passos mencionados anteriormente são utilizados como parte do programa Java no *framework* desenvolvido.

## 5.2 Criação do modelo

Como foi mencionado no Capítulo 3, para realizar a tarefa de classificação por meio de algoritmos supervisionado, precisamos de um *training set*, o nosso esteve formado por 6,687 *tweets* os quais foram rotulados manualmente.

As quantidades de amostras utilizadas para o classificador *Bullying Trace* foram de 4,788 *tweets rotulados com a classe “yes\_trace”* e 1,899 *tweets com a classe “não\_trace”*. Durante a classificação manual dos *tweets* como “no\_trace”, encontramos algumas descobertas (Tabela 4.2) que poderiam ser utilizadas em trabalhos futuros porque são sinais de quando o texto não deve ser considerado como um episódio de *bullying*.

Tabela 4.2 – Descobertas das amostras de treinamento rotuladas como “não” para o Classificador “*Bullying Trace*”

Descobertas das amostras rotuladas como “não”	Nº amostras
O texto expressa uma ideia/conselho/pergunta/dado estatístico sobre o bullying	649
O texto expressa uma suposição (I bet, I guess, I think, sth is like sth, sth looks like sth)	129
O texto está em futuro	123
O texto expressa uma condição: "Se ... Então"	120
O texto expressa uma probabilidade	56
O autor afirma que não é vítima	53
O autor expressa desejo que ninguém sofra de bullying nem ele/ela	44
O texto relata um sono	40
O texto é ficção	27
O autor deseja o mal a alguém	22
O autor afirma não viu nenhum episódio de bullying	18
O autor afirma que não é agressor	3

As quantidades de amostras, sendo em total 4,788 *tweets*, utilizadas para o classificador *Author Role* estão apresentadas na Figura 4.7, se observa que a maioria de amostras pertencem às classes “*reporter*” e “*victim*”

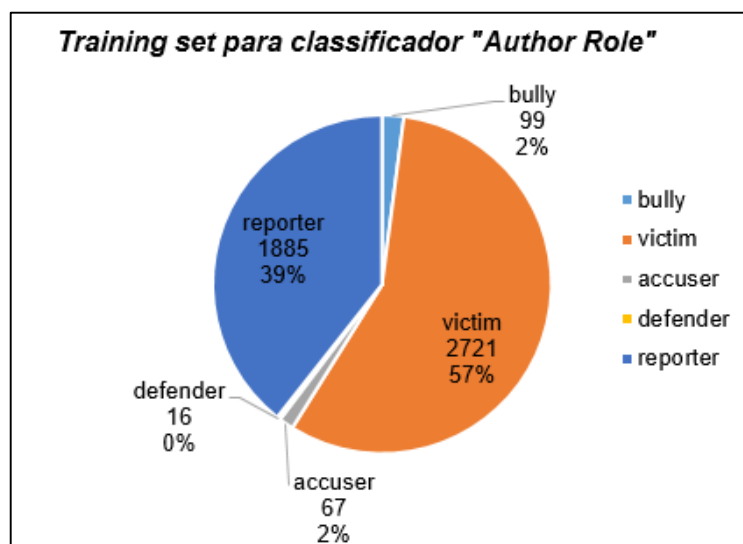


Figura 4.7– Número de amostras de treinamento para o Classificador “*Author Role*”

Existiram algumas descobertas durante a classificação manual como a existência de 34 amostras como *bully-victim* (textos escritos por agressores que se tornaram vítimas ou vice-versa), e 10 casos de autores da classe “*bully*” em que se

mostraram arrependido de serem agressores. Além disso na Tabela 4,3, se mostram as descobertas sobre as razões de porquê os autores se tornam vítimas de *bullying*.

Tabela 4.3 – Descobertas das amostras de treinamento rotuladas como “vítima” para o Classificador “*Author Role*”

Descobertas das amostras rotuladas como “vítima”	Nº amostras
Por aparência física	124
Por preferencias pessoais (música, livros, jogos, etc)	44
Por sua orientação sexual	29
Por ser ela/ele mesmo	19
Por pertencer a grupos étnicos	18
Por ser considerado como "diferente"	10
Por sofrer doenças	4
Por as características de sua família	4
Por pertencer a grupos religiosos	4

As quantidades de amostras, sendo em total 4,788 *tweets*, utilizadas para o classificador *Bullying Form* estão apresentadas na Figura 4.8, se observa que a maioria de amostras pertencem a classe “General” (92% do total). Entre as descobertas durante a classificação manual, se observou que o tipo general poderia ser subdivido em tipo “social” (exclusões, rumores, piadas) e “sexual” (assédio sexual).

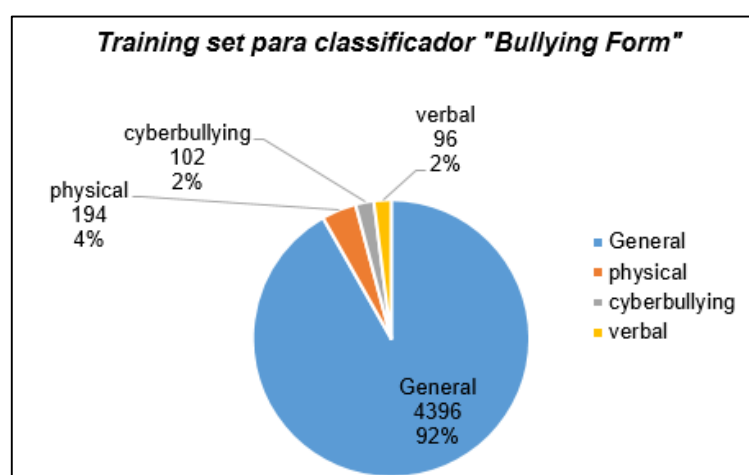


Figura 4.8– Número de amostras de treinamento para o Classificador “*Bullying Form*”

Não foi criado o modelo para o classificador “*Teasing Trace*” porque o número de amostras para a classe “yes” (que significa que o texto é só uma piada entre amigos ou sem severidade) era muito pequena comparada com a classe “não”.

Finalmente as amostras passaram pelo processo de *tokenization* da secção anterior e colocadas em um arquivo com extensão \*.arff (formato para WEKA). Na Figura 4.9, se mostra o arquivo \*.arff utilizado para o classificador *Bullying Trace*.

```
@relation bullying_tweets

@attribute text string
@attribute bullyingtrace {yes_trace,no_trace}

@data

'landons getting bullied at school already and im crying lmao',yes_trace
'#askdemi did you get bullied in school and if so how did you overcome it',no_trace
'these girls are being cyber bullied and harassed this account is claiming you agree t
'#askjacob whould u ever come to my school and talk about bullying since i get bullied
'#askkahlen wee you bullied in school ? and if so how did you handle it ? need some tip
'#asklilly how do you deal with bullying ? im being bullied and idk what to do about i
```

Figura 4.9– Arquivo \*.arff com amostras de treinamento para o Classificador “*Bullying Trace*”

Na Figura 4.10, apresenta-se o fluxo para a criação do modelo SVM com o pacote LIBLINEAR no *software* Weka, o mesmo fluxo foi utilizado para o pacote LibSVM e a árvore de decisão J48.

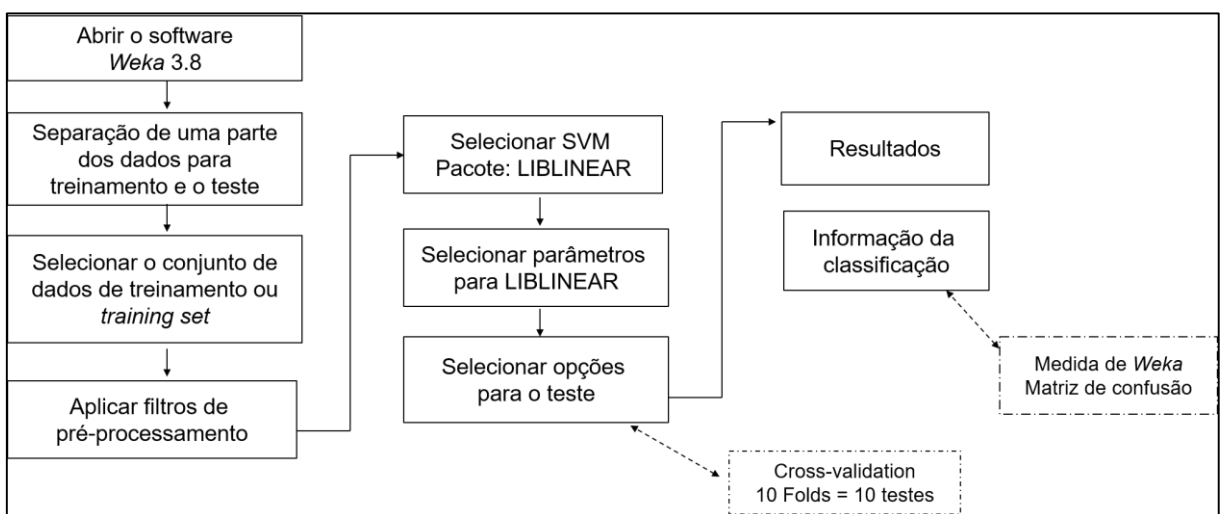


Figura 4.10 Fluxo para a criação de modelo SVM no *software* WEKA

Os resultados em detalhe da avaliação dos modelos estão disponíveis no Apêndice B. Como os resultados obtidos para os classificadores SVM foram muito similares aos desenvolvidos em (XU, 2015), não foram utilizados no framework mas permitiu conferir sua precisão.

### 5.3 Execução dos algoritmos SVM

Em (XU, 2015), foram desenvolvidos 5 classificadores SVM com o pacote LIBLINEAR e os modelos gerados foram utilizados por meio de um programa na linguagem Java. Para o projeto, utilizaram-se 4 dos 5 classificadores de (XU, 2015) e que são apresentados na Tabela 3.3. Cada classificador SVM possui seu próprio modelo em que encontra-se um peso calculado para cada *feature* do arquivo vocabulário. Adicionalmente, na Tabela 3.4 se mostram as acurácias e algumas observações dos resultados obtidos por os classificadores SVM selecionados.

Tabela 3.3 - Descrição dos classificadores SVM desenvolvidos em (XU, 2015)

<b>Classificador SVM</b>	<b>Descrição</b>	<b>Nº classes</b>	<b>Classes</b>
<i>Bullying Trace</i>	Permite identificar se o <i>tweet</i> faz alguma referência de existência de um episódio bullying.	2 (binário)	Yes, No
<i>Teasing Trace</i>	Permite identificar se o <i>tweet</i> possui falta de severidade ou seja se é uma possível piada entre amigos.	2 (binário)	Yes, No
<i>Author Role</i>	Permite identificar os papéis dos envolvidos em um episódio de <i>bullying</i>	6 (multi classe)	<i>Victim, Defender, Reporter, Bully, Accuser, Other</i>
<i>Bullying Form</i>	Permite identificar o tipo de bullying apresentado nos <i>tweets</i> .	4 (multi classe)	<i>General</i> (sem informação explícita o tipo de bullying), <i>Cyberbullying, Physical, Verbal</i>

Tabela 3.4 – Acurácia dos classificadores SVM desenvolvidos em (XU, 2015)

Classificador SVM	Acurácia (%)	Observações
<i>Bullying Trace</i>	86%	30.1%(9,764,583) dos <i>tweets</i> foram classificados como “ <i>bullying trace</i> ” de um total de 32,477,558 <i>tweets</i>
<i>Teasing Trace</i>	89%	A metade dos textos da classe “yes” não foram classificados corretamente. As possíveis justificativas são que os textos que podem ser piadas nem sempre estão acompanhados por <i>emoticons</i> ou <i>tokens</i> que expressem diversão. Por tanto, provavelmente a acurácia melhoraria com uma profundo análise de NLP ou com um maior número de amostras de classe “yes”.
<i>Author Role</i>	61%	Os papéis dos envolvidos <i>Accuser</i> (A), <i>bully</i> (B), <i>reporter</i> (R) e <i>Victim</i> (V) são os mais frequentes nas redes sociais. A maioria das amostras das classes R e V foram classificadas corretamente. Embora, não aconteceu o mesmo com as amostras das classes B e A. A justificativa é que provavelmente precisa-se de uma melhor representação das <i>features</i> para melhorar a predição.
<i>Bullying Form</i>	70%	O tipo de bullying “ <i>General</i> ” esteve presente na maioria das amostras - 95.2% (n = 9,296,651) classificadas como “ <i>bullying trace</i> ”. O segundo tipo de bullying mais frequente foi o <i>cyberbullying</i> (4.1%, n = 404,383 posts). Devido à pouca quantidade de exemplos para algumas classes, os classificadores não foram capazes de reconhecê-las corretamente.

Durante a análise dos algoritmos SVM de (XU, 2015), percebeu-se que os classificadores SVM devolvem uma classe categórica (rótulo) associada a um número real. Para os casos de os classificadores SVM binários, as amostras rotuladas como “No” possuem valores menores ou igual a zero, enquanto as amostras rotuladas como “Yes” possuem valores maiores a zero. Para os classificadores SVM multi-classe, um número real é calculada por cada classe, e a resposta final é o maior número dentro esses números calculados.

No projeto, se realizaram algumas modificações ao código Java: (1) criação de método para ler os textos coletados por meio da rede social “*Class21*” (2) definição de uma ordem de execução dos classificadores SVM, que consistiu em executar primeiro o classificador *Bullying Trace* como primeiro filtro, porque só os textos classificados como “yes” continuaram sendo avaliados por os outros três classificadores selecionados: *Teasing Trace*, *Author Role* e *Bullying Form*.

Por exemplo, o texto “*I was bullied in elementary school because of my height and then teased for crying about it I didnt have close friends until 3rd grade*”, passou o filtro porque trata-se de um episódio de bullying (classe “yes” do classificador *Bullying Trace*). Logo, os outros classificadores deram como resposta as classes “general”



(*Bullying Form*), “No” (*Teasing Trace*) e “Victim” (*Author role*). Observou-se que as classes estão associadas a um valor numérico real e variam em um intervalo que depende do classificador SVM.

## 5.4 Fuzzy Logic System para Bullying

A metodologia utilizada para a criação do Sistema de Lógica Fuzzy foi adaptada da metodologia proposta em (EMAMI; TÜRKSEN; GOLDENBERG, 1998). Consistiu em duas partes: (1) Identificação da estrutura e (2) Identificação dos parâmetros.

### 5.4.1 Identificação da estrutura

Este passo está composto por (1) a seleção das variáveis de entrada mais importantes assim como a definição de suas funções de pertinência (2) especificação das relações entre as variáveis de entrada e saída por meio de regras.

#### ***Input selection and membership***

Quando todos os textos são classificados com seus rótulos por meio da execução dos algoritmos SVM, foram analisados pelos autores do projeto. Selecionamos oito classes como variáveis de entrada: as classes “*victim*” e “*bully*”, duas das seis classes disponíveis para o autor do texto; as quatro classes disponíveis para o tipo de bullying como são “*general*”, “*physical*”, “*verbal*”, “*cyberbullying*”; e a variável “*teasing*”.

Para o caso do classificador de “*Author Role*” não foram utilizadas as classes “*Assistant*”, “*Accuser*”, “*Defender*” porque a precisão dos algoritmos para detectar essas classes foi muito baixa devido à falta de amostras para o treinamento e sua complexidade para diferenciá-los, mas a classe “*Reporter*” poderia ser utilizada para trabalhos futuros porque era uma das classes com maior quantidade de amostras. As classes do classificador de “*Bullying Form*” foram consideradas todas as classes por sua importância e gravidade como é o *bullying* físico que tem muitas consequências negativas para as vítimas. A variável de *teasing* embora não possua uma boa precisão

ela foi utilizada porque permitiria diminuir o nível de severidade do texto se ele fosse considerado como uma “piada” ou uma forma divertida entre amigos.

Logo, inicia-se o processo de *Fuzzificação*, que significa definir as funções de pertinência às variáveis selecionadas anteriormente. Cada variável de entrada associada ao tipo de bullying (“*Bullying Form*”) e ao autor (“*Author Role*”) possuem três funções de pertinência: *LOW*, *MEDIUM* e *HIGH*. A variável “*teasing*” possui duas funções de pertinência: “*YES*” e “*NO*”. Os intervalos de cada função de pertinência foi ajustado baseados na análise do conjunto de teste (*test set*) formado por 18,400 *tweets* quando eles foram avaliados só por os classificadores SVM.

A variável de saída e o nível de severidade “*severity*” e possui três funções de pertinência: *MODERATE*, *MAJOR* e *SEVERE*, esses nomes foram extraídos da *Bullying Matrix Assessment* (MELHUIH; PALMER; PASLEY, 2015) explicada no Capítulo 2.

### **Geração de regras**

Baseado na revisão da literatura do *bullying* do Capítulo 2, os autores do projeto definimos os critérios para a criação das regras. Por exemplo, avaliar a importância de cada variável de entrada para determinar a variável de saída “*severity*”. De fato, a variável de entrada “*physical*” tinha maior peso ou importância sobre as outras variáveis que são “*verbal*”, “*general*” e “*cyberbullying*”. Com respeito a variável autor, as classes “*victim*” e “*bully*” foram consideradas como importantes porque são as que apresentam maiores consequências negativas.

Finalmente, a variável “*teasing*”, foi considerada que quando pertence à função de pertinência “*no*” ela possui maior importância porque aumenta a severidade do episódio. Foram criados oito conjunto de regras modeladas no software *MATLAB* com o *Mamdani* como sistema de inferência *Fuzzy*:

- *Victim\_General*: Este conjunto de regras é utilizado quando o texto é classificado pelo classificador “*Author Role*” como “*victim*” e pelo classificador “*Bullying Form*” como “*General*”.
- *Victim\_Physical*: Este conjunto de regras é utilizada quando o texto é classificado pelo classificador “*Author Role*” como “*victim*” e pelo classificador “*Bullying Form*” como “*Physical*”.
- *Victim\_Verbal*: Este conjunto de regras é utilizada quando o texto é classificado pelo classificador “*Author Role*” como “*victim*” e pelo classificador “*Bullying Form*” como “*Verbal*”

- **Victim\_Cyberbullying:** Este conjunto de regras é utilizada quando o texto é classificado pelo classificador “*Author Role*” como “*victim*” e pelo classificador “*Bullying Form*” como “*Cyberbullying*”
- **Bully\_General:** Este conjunto de regras é utilizado quando o texto é classificado pelo classificador “*Author Role*” como “*bully*” e pelo classificador “*Bullying Form*” como “*General*”.
- **Bully\_Physical:** Este conjunto de regras é utilizada quando o texto é classificado pelo classificador “*Author Role*” como “*bully*” e pelo classificador “*Bullying Form*” como “*Physical*”.
- **Bully\_Verbal:** Este conjunto de regras é utilizada quando o texto é classificado pelo classificador “*Author Role*” como “*bully*” e pelo classificador “*Bullying Form*” como “*Verbal*”
- **Bully\_Cyberbullying** Este conjunto de regras é utilizada quando o texto é classificado pelo classificador “*Author Role*” como “*bully*” e pelo classificador “*Bullying Form*” como “*Cyberbullying*”

Cada conjunto de regras possui em total 18 regras. Na tabela 4.4, se apresentam as regras para o conjunto Victim\_General, por exemplo, para esse conjunto a regra 1 se interpreta desta forma:

```
IF general IS Low AND victim IS Low AND teasing IS Yes
THEN Severity IS Moderate
```

Tabela 4.4 – Regras para as variáveis de “Victim” e “General”

Regra	General	Victim	Teasing	Severity	Regra	General	Victim	Teasing	Severity
1	Low	Low	Yes	Moderate	10	Medium	Medium	No	Major
2	Low	Low	No	Moderate	11	Medium	High	Yes	Major
3	Low	Medium	Yes	Moderate	12	Medium	High	No	Major
4	Low	Medium	No	Moderate	13	High	Low	Yes	Major
5	Low	High	Yes	Moderate	14	High	Low	No	Major
6	Low	High	No	Moderate	15	High	Medium	Yes	Major
7	Medium	Low	Yes	Moderate	16	High	Medium	No	Severe
8	Medium	Low	No	Moderate	17	High	High	Yes	Major
9	Medium	Medium	Yes	Moderate	18	High	High	No	Severe

## 5.4.2 Identificação de parâmetros

### *Afinação das funções de pertinência*

A análise dos *tweets* do conjunto de teste permitiu escolher a forma da função de pertinência tanto para as variáveis de entrada como a de saída. O intervalo de cada

variável de entrada foi determinada pelo máximo e mínimo valor associada às classes selecionada no conjunto de teste de 18,400 *tweets*. Por outro lado, o intervalo da variável de saída “severidade” foi determinada pelo intervalo definido na *Bullying Matrix Assessment* (MELHUIH; PALMER; PASLEY, 2015) (8-9, 6-7 ou 3-5 para *MODERATE*, *MAJOR*, *SEVERE* respectivamente.)

### Ajustes de parâmetros

Em (SIVANANDAM; SUMATHI; DEEPA, 2007), mencionam-se sete métodos para o processo de *Defuzzificação*, sendo o método de “Centro de Gravidade” o amplamente usado. Portanto foi o que decidimos usar também por sua precisão.

Na Figura 4.11, se apresenta um exemplo prático do uso do sistema de Lógica Fuzzy, temos o texto “*I was bullied in elementary school because of my height and then teased for crying about it. I didn’t have close friends*”, o qual foi classificado como autor=“victim”, teasing=“yes” e bullyin form=“general”. Portanto, se utilizou o conjunto de regras “Victim\_General” o que avaliou que o texto pertence a função de pertinência de “Severe” como seu nível de severidade

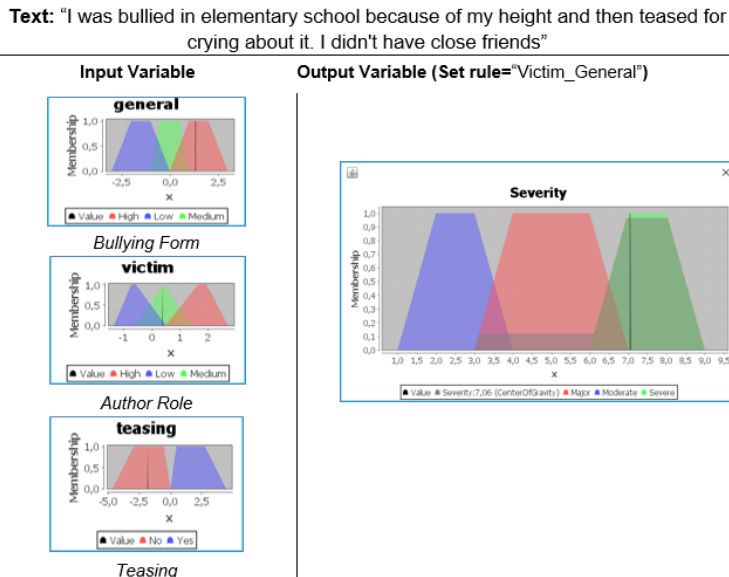


Figura 4.11 Exemplo de avaliação de um texto utilizando Lógica Fuzzy

## 5.5 Desenvolvimento da rede social interna “Class21” e o Sistema em *Java Swing Application* para a aplicação do *framework*

No projeto desenvolveu-se a rede social interna “Class21” para coletar novos textos de 140 caracteres, se manteve esse limite porque em (XU, 2015) menciona-se que os resultados dos algoritmos para textos maiores a 140 caracteres não é garantido. O “Class21” (Figura 4.12) foi desenvolvida em *Wordpress*<sup>11</sup> que é um CMS (Sistema de Gerenciamento de Conteúdo) para web e utiliza a base de dados *MySQL*.

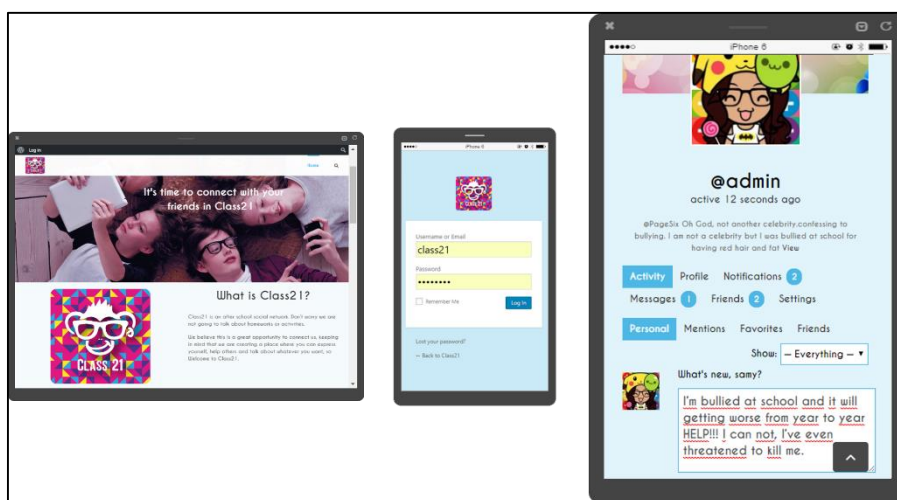


Figura 4.12 Rede Social “Class21”

O Sistema desenvolvido em *Java Swing* permitiu integrar os classificadores *SVM* e o sistema de Lógica *Fuzzy* (FLS). O FLS foi desenvolvido por meio da biblioteca *jFuzzyLogic.jar*. Na Figura 4.13, se apresenta o sistema em *Java Swing*, ele poderia ser utilizado pelo pessoa da escola para avaliar os textos coletados pelos seus alunos quando utilizarem o “Class21” ou também alguma outra ferramenta para coleta de dados.

<sup>11</sup> Wordpress - <https://es.wordpress.com>

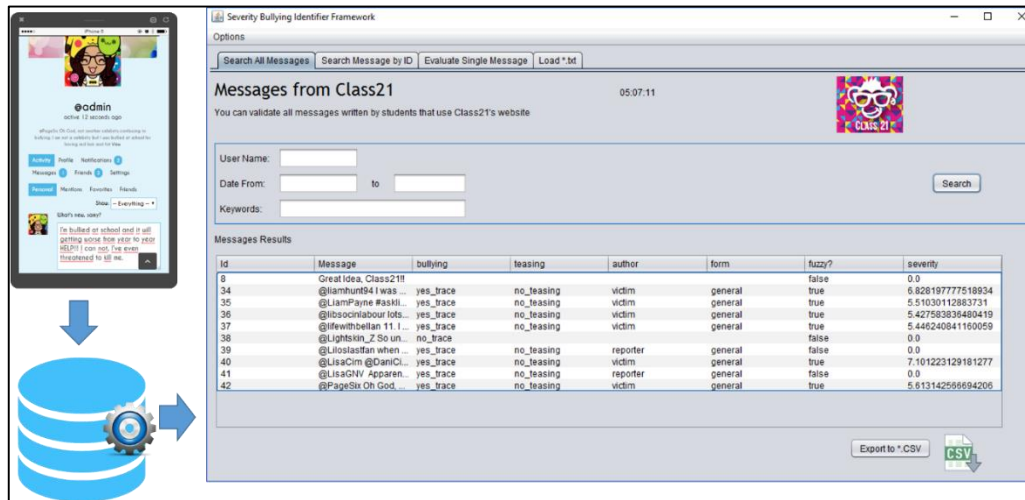


Figura 4.13 Software em Java para execução do *framework*

## 5.6 Resultados

Na Figura 4.14, se apresentam como foi a avaliação dos 18,400 *tweets* por meio do *framework*. Do total, unicamente 5,669 *tweets* possuem as variáveis de interesse e portanto passaram pelo sistema de *Lógica Fuzzy*

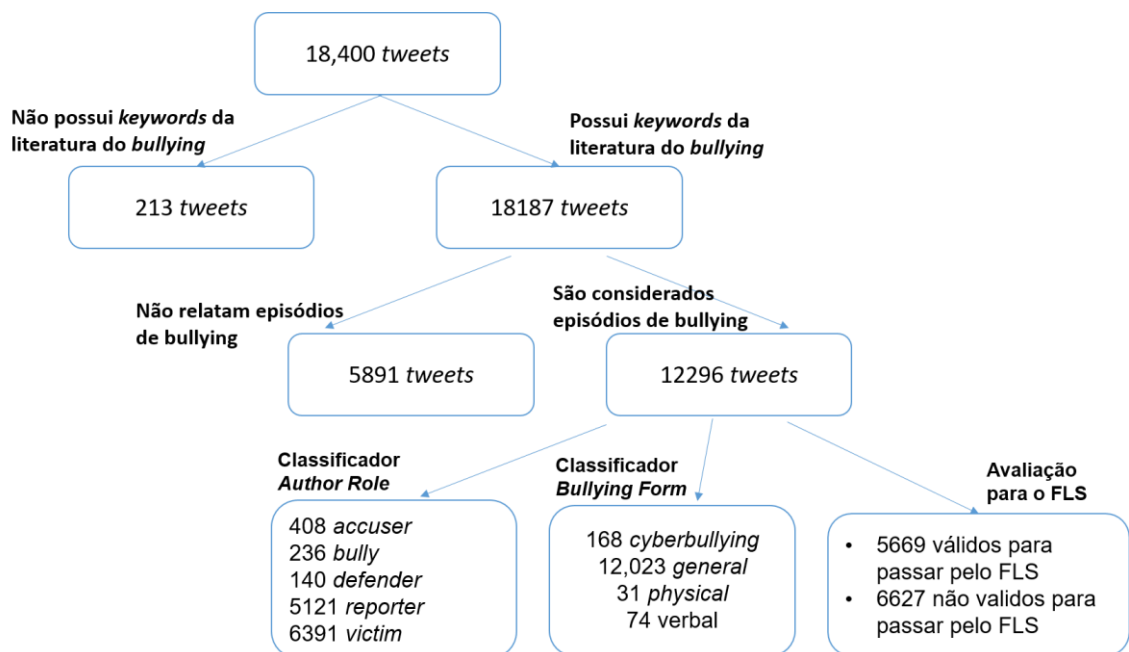


Figura 4.14 Classificação dos *tweets* por meio do *framework*.

Alguns dos *tweets* avaliados se apresentam na Tabela 4.5.

Tabela 4.5 – Exemplo de Tweets avaliados no *framework*

<b>Tweet</b>	<b>Severidade</b>
primary and middle school were such hard times for me i was bullied so much	7.5
Cyber bullying kills me on fb. I'm deleting it. You guys understand me and actually like me for who I am on here.	7.5
.@NathanGambleNG I get bullied in school for not having one please I beg of you please help me you're my only hope and chance	7.5
@the_author_ I had been bullied at school and my mother still doesn't understand why would kids treat me badly if I did nothing.	5.848829121
I got bullied a lot during my school years and it gave me really bad anxiety but I'm trying to overcome it and not let it effect me forever	5.800278262
Most likely because I was bullied in school and called ugly and the first girls to give me attention were dark skinned girls	2.607171783

## 5.7 Conclusões

O *bullying* é um problema muito sério que afeta a todas escolas no mundo, sem nenhuma distinção, todas as crianças estão expostas a serem vítimas ou se tornar agressoras se as escolas, os pais e a políticas de governo não desenvolvem programas anti-*bullying* que permitam a detecção e prevenção dos episódio de bullying que na maioria casos acontece no entorno escolar. A revisão da literatura de *bullying* permitiu entender os principais aspectos do problema e tentar desenvolver soluções tecnologias que ajudem a combater este problema.

No Apêndice A, se avaliaram muitos fatores que podem causar *bullying* e que também foram conferidos durante a fase de treinamento enquanto se realizou a classificação manual (colocar rótulos a cada *tweet* do *training set*). A criação dos modelos mostraram que os SVM utilizando o pacote LIBLINEAR obtiveram melhores resultados e a criação do modelo assim como sua execução foi de maior velocidade que os algoritmos do pacote LibSVM e a árvore de decisão J48. O LIBLINEAR trabalha com eficiência para os problemas de tipo linear, sem *kernel*, é por isso sua rapidez

para avaliar dados (no nosso caso os tweets) com uma grande quantidade de *features* (número de índices utilizados na representação dos dados no formato de vetor).

As técnicas supervisionadas de Machine Learning demonstraram que quando se possui uma grande quantidade de dados treinamento, elas conseguem melhorar a precisão, além disso é importante que o *training set* esteja balanceado, o que significa que as classes possuem o mesmo número de amostras.

A precisão dos algoritmos SVM afetam diretamente o sistema de Lógica *Fuzzy*, se as classes selecionadas para os textos não são corretos, o nível de severidade obtido pelas regras *Fuzzy* não será o adequado. A avaliação da severidade obtida pelo FLS foi realizada de acordo à percepção da autora deste projeto e os professores envolvidos, mas a avaliação por parte de especialistas, como psicólogos, ajudaria a determinar com maior precisão se os textos possuem os níveis de *bullying* corretos.

Por outro lado, o desenvolvimento da rede social *Class21* e do sistema em *Java Swing* permite de maneira prática executar os algoritmos, é assim que o pessoal da escola, se utilizar o framework, conseguiria avaliar de forma mais rápida e efetiva os textos escritos por seus alunos.

## 5.8 Trabalhos Futuros

O número de amostras utilizadas para o treinamento dos algoritmos SVM (6,687 *tweets*) tentou-se aproximar ao número utilizado no trabalho de (XU, 2015) que foi de 7,321 *tweets*, embora, poderia se estabelecer um número de amostras com maior significado por meio de métodos de amostragem.

A criação dos novos modelos para a classificação permitiu conferir os resultados dos modelos utilizados neste projeto. Contudo esses novos não foram adaptados ao *framework* porque os resultados obtidos durante a etapa de treinamento não foram muito melhores dos achados na literatura. Deve-se considerar melhorar os classificadores para determinar ao autor do texto especialmente para a classe de “Agressor”, porque de acordo com a literatura, as crianças envolvidas em episódios de *bullying* com um padrão de agressor também sofrem e precisam de ajuda que em



alguns casos pode ser muito diferenciada da ajuda para as vítimas. O classificador de *teasing* foi quem deu taxas de precisão muito baixas (menores do que 50%) para a classe “yes”, portanto é necessário melhorar a forma de representar os *tweets* e provavelmente adicionar técnicas de análise de sentimento. A precisão para o classificador de *bullying form* para a classe “physical” (bullying físico) poderia ser melhorado também com outras técnicas e achamos que é prioridade esse classificador porque o *bullying* físico tem muito impacto negativo para as vítimas.

Os algoritmos SVM foram desenvolvidos para a classificação dos textos na língua inglesa mas poderia se utilizar o mesmo processo de desenvolvimento para a língua portuguesa. A coleta de *tweets* em português, para treinamento e teste, é possível por meio do *Python* e a biblioteca do *Twitter*. Contudo, a precisão da classificação vai depender das técnicas mais apropriadas a serem utilizadas para a limpeza e transformação dos textos em vetores para a gramática portuguesa, que não necessariamente podem ser as mesmas utilizadas neste projeto.

Durante a fase de colocar as classes de forma manual aos *tweets* para serem parte do conjunto de treinamento, descobriu-se que as vítimas expressam seu sofrimento e relatam o que sentem ou porque as razões de porquê são agredidas. Por outro lado, também descobriu-se razões de porquê o texto não deveria ser considerado como episódio de *bullying* como por exemplo textos que falam no tempo futuro, com condições “Se .. *então*” ou falam de episódios de *bullying* fictícios (que pertencem a livros, filmes, música, etc.) . Essas descobertas poderiam ser utilizadas para melhorar os classificadores, para o caso do classificador de “*Bullying Trace*”, o uso de técnicas de *Named Entity Recognizer* (NER<sup>12</sup>) permitiria detectar se as pessoas ou as coisas escritas no textos são fictícias ou não. Outra possibilidade é que utilizando técnicas de agrupamento (algoritmos não supervisionados) se consiga subdividir aos autores ou os tipos de *bullying*.

Um grupo de autores que não foi considerado são os que pertencem a “*bully-victim*”, este grupo na literatura do *bullying* se indica que é muito complexo mas não é pouco frequente, porque tem casos em que as vítimas se tornam agressoras. Portanto, deve-se desenvolver um classificador para identificar esse grupo de autores.

---

<sup>12</sup> Stanford Named Entity Recognizer (NER) - <https://nlp.stanford.edu/software/CRF-NER.shtml>

As regras para a Lógica *Fuzzy* foram definidos pelos autores deste projeto, recomenda-se que elas sejam definidas por especialistas no tema de bullying como os *psicólogos*. Além disso, deve-se testar o *framework* em um entorno real, ou seja uma turma de escola para assim avaliar a efetividade real do *framework*. Neste ponto, os novos textos para coletar os dados poderiam ser a rede social “Class21” ou por meio de jogos de tipo *Serius Games* porque eles se adaptam aos comportamentos dos jogadores e isso permitiria avaliar suas reações naturais diante de desafios, além como foi apresentado na literatura, as crianças quando não se sentem supervisionadas, reagem de forma natural.

# REFERÊNCIAS

---

---

AFONSO MAZON, J. **Projeto de estudo sobre ações discriminatórias no âmbito escolar, organizadas de acordo com áreas temáticas, a saber, étnico-racial, gênero, geracional, territorial, necessidades especiais, socioeconômica e orientação sexual**, 2009. .

AJAO, O.; HONG, J.; LIU, W. A survey of location inference techniques on Twitter. **Journal of Information Science**, v. 41, n. 6, p. 855–864, 1 dez. 2015.

ANTONIO, A.; NETO, L.; FILHO, L. M.; SAAVEDRA, L. H. Programa de redução do comportamento agressivo entre estudantes. **Associação Brasileira de Multiprofissional de Proteção à Infância e à Adolescência**, 2003.

BAI, Y.; WANG, D. Fundamentals of Fuzzy Logic Control – Fuzzy Sets, Fuzzy Rules and Defuzzifications. In: BAI, Y.; ZHUANG, H.; WANG, D. (Ed.). **Advanced Fuzzy Logic Technologies in Industrial Applications**. 1. ed. [s.l.] Springer-Verlag London, 2007. p. 17–36.

BARTON, E. A. The Bully, Victim, and Witness Relationship Defined. In: **Bully prevention: Tips and strategies for school leaders and classroom teachers**. 2nd. ed. [s.l.: s.n.]p. 1–18.

BAUER, N. S.; LOZANO, P.; RIVARA, F. P. The effectiveness of the Olweus Bullying Prevention Program in public middle schools: a controlled trial. **The Journal of adolescent health**, v. 40, n. 3, p. 266–274, mar. 2007.

BELLMORE, A.; CALVIN, A. J.; XU, J. M.; ZHU, X. The five W's of “bullying” on Twitter: Who, What, Why, Where, and When. **Computers in Human Behavior**, v. 44, p. 305–314, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.chb.2014.11.052>>.

BEUREN, I. M. **Como elaborar trabalhos monográficos em contabilidade - teoria e prática**. 3. ed. [s.l.] Atlas, 2015.

BOMMERSON, B. Machine learning to classify bullying messages on Twitter Research questions. n. 2005, p. 1–14, 2015.

CENTERS FOR DISEASE CONTROL. **The Relationship Between Bullying and Suicide : What We Know and What it Means for Schools**National Center for Injury Prevention and Control, 2014. .

CHANG, C.; LIN, C. LIBSVM : A Library for Support Vector Machines. p. 1–39, 2013.

CLÁUDIO, A. P.; CARMO, M. B.; BIOSYSTEMS, B.; CARVALHOSA, S.; CANDEIAS, M. D. J.; GASPAR, A. A serious game-based solution to prevent bullying. 2015.

CRAMMER, K.; SINGER, Y. On the Learnability and Design of Output Codes for Multiclass Problems. **Machine Learning**, v. 47, n. 2, p. 201–233, maio 2002.

CRICK, N. R.; GROTPETER, J. K. Children's treatment by peers: Victims of relational and overt aggression. **Development and Psychopathology**, v. 8, n. 2, p. 367, 4 mar. 1996.

DAKE, JOSEPH A; PRICE, JAMES H; TELLJOHANN, S. K. The nature and extent of bullying at school. **The Journal of School Health**, v. 73, n. May, p. 173–180, 2003.

DINAKAR, K.; JONES, B.; HAVASI, C.; LIEBERMAN, H.; PICARD, R. Common Sense Reasoning for Detection, Prevention, and Mitigation of Cyberbullying. **ACM Transactions on Interactive Intelligent Systems**, v. 2, n. 3, p. 1–30, 1 set. 2012. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2362394.2362400>>. Acesso em: 11 maio. 2014.

DINAKAR, K.; REICHART, R.; LIEBERMAN, H. Modeling the Detection of Textual Cyberbullying. **Association for the Advancement of Artificial Intelligence**, p. 11–17, 2011. Disponível em: <<http://www.cl.cam.ac.uk/~rr439/papers/3841-16937-1-PB.pdf>>.

DULMUS, C. N.; SOWERS, K. M.; THERIOT, M. T. Prevalence and Bullying Experiences of Victims and Victims Who Become Bullies (Bully-Victims) at Rural Schools. **Victims & Offenders**, v. 1, n. 1, p. 15–31, 2006.

EMAMI, M. R.; TÜRKSEN, I. B.; GOLDENBERG, A. A. Development of a systematic methodology of fuzzy logic modeling. **IEEE Transactions on Fuzzy Systems**, v. 6, n. 3, p. 346–361, 1998.

EVANS, C. B. R.; FRASER, M. W.; COTTER, K. L. The effectiveness of school-based bullying prevention programs: A systematic review. **Aggression and Violent Behavior**, v. 19, n. 5, p. 532–544, 2014.

EVANS, C. B. R.; SMOKOWSKI, P. R. Understanding weaknesses in bullying research: How school personnel can help strengthen bullying research and practice. **Children and Youth Services Review**, v. 69, n. August, p. 143–150, 2016.

FAN, R.-E.; CHANG, K.-W.; HSIEH, C.-J.; WANG, X.-R.; LIN, C.-J. LIBLINEAR: A Library for Large Linear Classification. **Journal of Machine Learning Research**, v. 9, n. 2008, p. 1871–1874, 2015.

FERNÁNDEZ-GAVILANES, M.; ÁLVAREZ-LÓPEZ, T.; JUNCAL-MARTÍNEZ, J.; COSTA-MONTENEGRO, E.; JAVIER GONZÁLEZ-CASTAÑO, F. Unsupervised method for sentiment analysis in online texts. **Expert Systems with Applications**, v. 58, p. 57–75, 2016.

FISSETTE, M. Author identification in short texts. 2010.

GLADDEN, R. M.; VIVOLO-KANTOR, A. M.; HAMBURGER, M. E.; LUMPKIN, C. D. **Bullying surveillance among youths** Centers for Disease Control and Prevention Atlanta, Georgia, 2014. .

GOLDWEBER, A.; WAASDORP, T. E.; BRADSHAW, C. P. Examining the link between forms of bullying behaviors and perceptions of safety and belonging among secondary school students. **Journal of School Psychology**, v. 51, n. 4, p. 469–485, 2013.

GOURNEAU, B. Students' Perspectives Of Bullying In Schools. v. 5, n. 2, p. 117–127, 2012.

GUGGENBERGER, A. Another Introduction to Support Vector Machines. 2012.

GUYON, I.; ELISSEEFF, A. An Introduction to Feature Extraction. In: **Feature Extraction: Foundations and Applications**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 1–25.

HINDUJA, S.; PATCHIN, J. W. Bullying, Cyberbullying, and Suicide. **Archives of Suicide Research**, v. 14, n. 3, p. 206–221, 2010.

HSU, C.-W.; CHANG, C.-C.; LIN, C.-J. **A Practical Guide to Support Vector Classification**, fev. 2016. .

HSU, C.; LIN, C. A Comparison of Methods for Multi-class Support Vector Machines. **IEEE Transactions on Neural Networks**, v. 13, p. 415–425, 2002.

IANNOTTI, R. J. Health Behavior in School-Aged Children ( HBSC ), 2009-2010 Codebook : Student Survey. p. 2009–2010, 2010.

J.HAN, J.PEI, M. K. **Data Mining: Concepts and Techniques**, 2012. .

JAIN, a. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Computing Surveys**, v. 31, n. 3, p. 264–323, 1999.

JAVIER, R. A.; DILLON, J. Bullying and its consequences: In search of solutions. **Journal of Social Distress & the Homeless**, v. 22, n. 1, p. 1–6, 2013.

JURAFSKY, D.; MARTIN, J. H. N-Grams. In: **Speech and Language Processing**. [s.l: s.n.]p. 42–69.

JURAFSKY, D. S.; MARTIN, J. H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**, 2006. .

KONTOSTATHIS, A.; REYNOLDS, K.; GARRON, A.; EDWARDS, L. **Detecting Cyberbullying: Query Terms and Techniques** **Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13**, 2013. .

KOTSIANTIS, S. B. Supervised Machine Learning: A Review of Classification Techniques. **Emerging Artificial Intelligence Applications in Computer Engineering**, v. 31, p. 249–268, 2007.

LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, v. 33, n. 1, 1977.

LIDDY, E. D. **Natural Language Processing** **Encyclopedia of Library and Information Science**, 2001. .

LOPES NETO, A. A. Bullying: comportamento agressivo entre estudantes. **Jornal de Pediatria**, v. 81, n. 5, p. 164–172, nov. 2005.

LORENA, A. C.; DE CARVALHO, A. C. P. L. F. Uma Introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007.

MANCILLA-CACERES, J. F.; PU, W.; AMIR, E.; ESPELAGE, D. A Computer-in-the-Loop Approach for Detecting Bullies in the Classroom. **Proceedings of the 5th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction**, p. 139–146, 2012.

MELHUIISH, N.; PALMER, R.; PASLEY, S. **Bullying prevention and response : A guide for Schools**, 2015. .

MELO, J. H. F. C. H. de; CAVALCANTI, C. da R. S. M. T. **Lógica Fuzzy aplicada às Engenharias**, 2012. .

NALINI, K.; SHEELA, L. J. **Classification of Tweets Using Text Classifier to Detect Cyber Bullying** **Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2**, 2015. .  
Disponível em: <[http://link.springer.com/10.1007/978-3-319-13731-5\\_69](http://link.springer.com/10.1007/978-3-319-13731-5_69)>.

NANDHINI, B. S.; SHEEBA, J. I. Online Social Network Bullying Detection Using Intelligence Techniques. **Procedia Computer Science**, v. 45, p. 485–492,

2015.

OLWEUS, D. **Bullying in school: What we know and what we can do**, 1993.

OLWEUS, D. A Profile of Bullying at School. **Educational Leadership**, v. 60, n. 6, p. 12–17, 2003.

ONG, F. **Bullying at School**, 2003. .

PARIME, S.; SURI, V. Cyberbullying detection and prevention: Data mining and psychological perspective. **2014 International Conference on Circuits, Power and Computing Technologies, ICCPCT 2014**, p. 1541–1547, 2014.

PÉREZ, V. **Percepción de Gravedad, Empatía y Disposición a Intervenir en Situaciones de Bullying Físico, Verbal y Relacional en Profesores de 5° a 8° Básico**, 2011. .

PIOTROWSKI, D.; HOOT, J. Bullying and violence in schools: What teachers should know and do. **Childhood Education**, v. 84, p. 357–363, 2008.

PLATT, J. C. Advances in Kernel Methods. In: SCHÖLKOPF, B.; BURGESS, C. J. C.; SMOLA, A. J. (Ed.). Cambridge, MA, USA: MIT Press, 1999. p. 185–208.

RAGOZZINO, K.; UTNE O'BRIEN, M. **Social and Emotional Learning and Bullying Prevention**, 2009. .

RAMASUBRAMANIAN, C.; RAMYA, R. Effective Pre-Processing Activities in Text Mining using Improved Porter's Stemming Algorithm. **International Journal of Advanced Research in Computer and Communication Engineering**, v. 2, n. 12, p. 4536–4538, 2013.

REESE, R. **Natural language processing with java**. [s.l.] Packt Publishing, 2015.

RETTEW, D. C.; PAWLOWSKI, S. Bullying. **Child and Adolescent Psychiatric Clinics of North America**, v. 25, n. 2, p. 235–242, 2016.

REZENDE, S. O.; MARCACINI, R. M.; MOURA, M. F. O uso da Mineração de Textos para Extração e Organização Não Supervisionada de Conhecimento. **Revista de Sistemas de Informação da FSMA**, v. 7, p. 7–21, 2011.

RICHARDSON, D.; HIU, C. F. **Ending the torment: tackling bullying from the schoolyard to cyberspace**. [s.l.: s.n.]. Disponível em: <[http://srsg.violenceagainstchildren.org/sites/default/files/2016/Ending\\_bullying/bullyingreport.pdf](http://srsg.violenceagainstchildren.org/sites/default/files/2016/Ending_bullying/bullyingreport.pdf)>.

RICHARDSON, R. J. **Pesquisa social: métodos y técnicas**. [s.l.] Atlas, 2007.

ROSSI, M.; BENATTI, S.; FARELLA, E.; BENINI, L. Hybrid EMG classifier based on HMM and SVM for hand gesture recognition in prosthetics. **Proceedings of the IEEE International Conference on Industrial Technology**, v. 2015–June, n. June, p. 1700–1705, 2015.

SAINIO, M.; VEENSTRA, R.; HUIJSING, G.; SALMIVALLI, C. Same- and Other-Sex Victimization: Are the Risk Factors Similar? **Aggressive Behavior**, v. 38, n. 6, p. 442–455, 2012.

SHAHEEN, S. Cyber-Bullying: Issues and Solutions for the School, the Classroom and the Home. 2009.

SIVANANDAM, S. N.; SUMATHI, S.; DEEPA, S. N. **Introduction to fuzzy logic using matlab**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

TUSINSKI, K. **The Causes and Consequences of Bullying**, 2008. .

UNESCO. **School Violence and Bullying: Global Status Report**, 2017. .

VAN GOETHEM, A. a J.; SCHOLTE, R. H. J.; WIERS, R. W. Explicit- and implicit bullying attitudes in relation to bullying behavior. **Journal of abnormal child psychology**, v. 38, n. 6, p. 829–842, ago. 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20352324>>. Acesso em: 13 nov. 2013.

VAN NOORDEN, T. H. J.; HASELAGER, G. J. T.; CILLESSEN, A. H. N.; BUKOWSKI, W. M. Empathy and Involvement in Bullying in Children and Adolescents: A Systematic Review. **Journal of Youth and Adolescence**, v. 44, n. 3, p. 637–657, 2014.

VERMA, T. Tokenization and Filtering Process in RapidMiner. **International Journal of Applied Information Systems**, v. 7, n. 2, p. 16–18, 2014.

VIJAYARANI, S.; ILAMATHI, J.; NITHYA, M. Preprocessing Techniques for Text Mining - An Overview. In: 1, **Anais...**2015. Disponível em: <<http://www.ijcscn.com/Documents/Volumes/vol5issue1/ijcscn2015050102.pdf>>.

VIJAYARANI, S.; JANANI, R. Text Mining: open Source Tokenization Tools – An Analysis. **Advanced Computational Intelligence: An International Journal (ACII)**, v. 3, n. 1, p. 37–47, 2016. Disponível em: <<http://aircconline.com/acii/V3N1/3116acii04.pdf>>.

WANG, C. **A study of membership functions on mamdani-type fuzzy inference system for industrial decision-making**. 2015. Lehigh University, 2015.

WANG, J.; IANNOTTI, R. J.; NANSEL, T. R. School Bullying Among Adolescents in the United States: Physical, Verbal, Relational, and Cyber. **Journal of**



**Adolescent Health**, v. 45, n. 4, p. 368–375, out. 2009. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1054139X09001384>>.

WANG, L. **Support vector machines: theory and applications**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. v. 177

WANG, Z.; XUE, X. Multi-Class Support Vector Machine. In: MA, Y.; GUO, G. (Ed.). **Support Vector Machines Applications**. Cham: Springer International Publishing, 2014. p. 23–48.

WAZLAWICK, R. S. Uma Reflexão sobre a Pesquisa em Ciência da Computação à Luz da Classificação das Ciências e do Método Científico. **Revista de Sistemas de Informação da FSMA**, v. nº 6, p. 3–10, 2010.

WOLKE, D.; LEREYA, S. T. Long-term effects of bullying. **Archives of disease in childhood**, v. 100, n. 9, p. 879–85, 2015.

XU, J. **Understanding and fighting bullying with machine learning**. 2015. 2015.

XU, J.-M.; JUN, K.-S.; ZHU, X.; BELLMORE, A. Learning from Bullying Traces in Social Media. p. 656–666, 2012.

XU, J.-M.; ZHU, X.; BELLMORE, A. Fast learning for sentiment analysis on bullying. **Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining - WISDOM '12**, p. 1–6, 2012. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2346676.2346686>>.

# APÊNDICE A- MODELO ESTATÍSTICO DO TIPO LOGIT PARA AVALIAÇÃO DE POTENCIAL DE *BULLYING* NO ENSINO FUNDAMENTAL

---

---

## A1. Introdução

Para verificar se alunos estão sofrendo “*bullying*” é importante que se tenha métodos que possam antecipar o diagnóstico que o próprio aluno pode não estar percebendo ou que não tenha coragem de relatar o que está ocorrendo. Este modelo estatístico é um modelo inicial simples (que pode ser melhorado) e deve ser utilizado em conjunto com outras ferramentas, por exemplo, as baseadas em Inteligência Artificial (*Machine Learning*), em Gerência do Conhecimento, etc. Nesse tipo de assunto a aplicação de vários métodos vai melhorar a interpretação final dos resultados e, principalmente, ajudar na validação das conclusões (um modelo pode validar o outro e vice-versa).

## A2. Objetivo

A fonte deste relatório foi o projeto de “Estudo sobre ações discriminatórias no âmbito Escolar, organizadas de acordo com áreas temáticas, a saber, étnico-racial, gênero, geracional, territorial, necessidades especiais, socioeconômica e orientação sexual” feito pela Fundação Instituto de Pesquisas Econômicas (FIPE) (AFONSO MAZON, 2009) e é considerada uma pesquisa pioneira que tem como objetivo analisar de maneira global e coerente a incidência de preconceito e discriminação nas escolas

públicas para avaliação de ações globais para transformar a escola em ambiente essencial ao convívio com as diversidades.

Neste projeto foram definidas 7 Áreas temáticas de discriminação (*Gênero, Geracional, Deficiência, Orientação Sexual, Socioeconômica, étnico racial, territorial*) e 9 grupos sociais pesquisados (*pobres, negros, índios, ciganos, moradores de periferia/favela, moradores de áreas rurais, homossexuais, portadores de deficiência mental e portadores de deficiência física*). Além disso, este projeto foi aplicado a diferentes públicos alvo, porém, o público de nosso interesse é: *Estudantes de 7 ou 8 séries de Ensino fundamental regular*.

A pesquisa da FIPE foi aplicada EM 501 escolas, com um total de 15,087 alunos onde o 25.6% tinham até 14 anos (3,863 alunos). As características avaliadas dos respondentes são apresentadas na tabela A1.

Tabela A2.1– Características dos respondentes do questionário FIPE

<b>Características avaliadas</b>	<b>Resultados gerais</b>
Gênero e faixa etária	Sexo feminino e faixas etárias altas têm menor atitude preconceituosa.
Região do país	Respondentes das Regiões Sudeste e Sul têm percepções menos preconceituosas, enquanto aqueles das regiões Nordeste e Norte tem maior grau de preconceito.
Acesso a meios de informação	Quanto maior o acesso à informação menor a percepção preconceituosa
Cor, etnia	Negros, mulatos, pardos e brancos apresentaram os menores níveis de preconceito. Amarelos ou orientais têm percepções mais preconceituosas.
Religião	Pessoas sem religião têm maiores valores para os índices de preconceito e discriminação.
Modalidade de ensino	Alunos de ensino fundamental apresentam maiores níveis de preconceito e distância social
<i>Porte da escola</i>	<i>Não relevante</i>
<i>Localização da escola</i>	<i>Não relevante</i>

### **A3. Natureza e Instrumentos da pesquisa da FIPE**

A natureza da pesquisa foi a mensuração de crenças, atitudes e comportamentos comuns e específicos entre diferentes populações-alvo.

O instrumento de pesquisa foi um questionário por cada população alvo. Portanto, definiu-se um questionário específico para os alunos da penúltima série (7/8 anos) do Ensino Fundamental. O questionário esteve composto por assuntos como: Exposição a mídia por parte dos respondentes, Hábitos de lazer, Escala de distância social, Crenças e atitudes, Conhecimento de práticas discriminatórias (*bullying*), sociodemográficas e Escolares. O questionário reflete o universo cognitivo dos respondentes com maior facilidade para a sua compreensão. Na tabela A2, apresentam-se os três indicadores obtidos da transformação das respostas dos questionários e na Tabela A3 estão alguns exemplos das frases que expressam preconceito.

Tabela A3.1 – Indicadores de acordo às respostas do questionário FIPE

Que foi avaliado?	Como foi avaliado?	Possíveis respostas	Indicador
(1) As crenças, atitudes e os valores que expressam preconceito.	Utilizaram-se 83 frases associadas às 7 áreas temáticas.	(1) Discordo Muito (2) Discordo Pouco (3) Concordo Pouco (4) Concordo Muito	Respostas transformadas em um <b>índice percentual de concordância (IPC%)</b> . Um maior nível de concordância de preconceito significa um maior % de IPC
(2) Mensuração da distância social em relação às pessoas dos grupos sócias pesquisados. A predisposição do respondente a manter proximidade com um determinado grupo social.	Assinalar apenas a frase com a qual concorda com maior intensidade dentre as listadas.	Exemplo: <i>Mensuração da distância social para o grupo social "Negro"</i> : (1) Aceitaria minha filha casar/namorar/estudar com ele. (2) Aceitaria como colega de trabalho/aluno na escola.	<b>Índice percentual de distância social (IPCD%)</b> variando de 0 a 100. Uma maior aceitação ou proximidade ao grupo social significa um menor % de IPCD.
(3) Conhecimento de práticas discriminatórias na escola ( <i>Bullying</i> ): existência de situações de humilhação, agressão física, acusação injusta em função do pertencimento aos grupos sociais.	Escala ordinal de 3 pontos que mede o grau de conhecimento de c/respondente acerca da ocorrência de <i>bullying</i> .	(1) Nem vi, nem soube que aconteceu nesta escola (2) Não vi, mas soube que aconteceu nesta escola (3) Vi nesta escola	<b>Índice percentual de conhecimento de situações de Bullying na escola (IPCSB%)</b> : média dos valores medidos em c/frase variando entre 0 e 100. 100% indica que o respondente viu ocorrer na escola situações de bullying.

O resultado da aplicação do questionário foi que se reconhece que o preconceito é um elemento presente no ambiente das escolas públicas do país e que são os alunos os que apresentam níveis altos de preconceito.

Tabela A3.2 – Exemplos de frases que expressa preconceito

Frase	Tipo de Preconceito
-------	---------------------

Os brancos, em geral, são mais estudiosos que os negros	Étnico racial
O número de deficientes físicos na escola é muito pequeno para se preocupar com eles	Necessidades especiais (deficiência)
Pessoas homossexuais não são confiáveis	Orientação Sexual
O estudante que entra mais velho na escola tem mais dificuldade de aprender	Geracional
Os estudantes do campo são mais lentos para aprender	Territorial
Existem trabalhos que devem ser realizados apenas por homens	Gênero
Os estudantes pobres são mais revoltados	Socioeconômico

#### **A.4 Análise da Atitude Preconceituosa, Distância Social e Conhecimento de Situações de Discriminação (*bullying*) presenciadas para o público alvo**

O público alvo avaliado está composto por alunos de penúltimo ano de Ensino Fundamental. Segundo os resultados da pesquisa, entre os alunos, as maiores vítimas de humilhação, agressão física e acusações injustas na escola são os alunos pobres, negros e, principalmente, os homossexuais. A ocorrência de situações de *bullying* de acordo com o conhecimento dos respondentes, por sua vez, é mais baixa do que a atitude preconceituosa (IPC%) e do que a distância social observada entre os respondentes (IPCD%). Este resultado está de acordo com as expectativas, pois a ocorrência de *bullying* representa a efetiva materialização do preconceito e da discriminação, que são comportamentos latentes que podem ou não resultar em ações discriminatórias por meio do *bullying*. Cabe ressaltar, no entanto, que apesar de apresentar um índice de ocorrência mais baixo nas escolas pesquisadas, o *bullying* é uma realidade nas escolas.

A seguir, apresentam-se os resultados dos indicadores IPC%, IPCD% e IPCSB% para os alunos de penúltimo ano de Ensino Fundamental

##### **A4.1 Atitude Preconceituosa – IPC (%)**

Os alunos do ensino fundamental apresentam as maiores médias (Tabela A4.1) para o índice de concordância com as frases que expressam atitudes preconceituosas.

Tabela A4.1 Atitude Preconceituosa – IPC % - para Alunos de Ensino Fundamental

<b>Atitude Preconceituosa</b>	<b>IPC %</b>
Gênero	42% (p.e. Os alunos concordam em um 42% com as frases de preconceito relacionado ao gênero)
Geracional	41%
Deficiência	35%
Sócio Económico	29%
Orientação Sexual	28%
Étnico – Racial	27%
Territorial	25%

### **A4.2 Distância Social – IPCD (%)**

As maiores médias para o IPCD e, portanto, maior distância social, foram aqueles formados por respondentes de até 14 anos de idade morenos, pardos, brancos e amarelos/orientais do sexo masculino.

Tabela A4.2 Distância Social – IPCD % - para Alunos de Ensino Fundamental

<b>Distância Social</b>	<b>IPCD %</b>
Homossexual	82% (p.e. O aluno evita ter contato com uma pessoa homossexual em um 82%)
Deficiente Mental	80%
Cigano	80%
Deficiente físico	72%
Pobre	71%
Índio	71%
Morador de periferia/favela	70%
Negro	65%
Morador de área rural	64%

### **A4.3 Conhecimento de situações de discriminação – IPCSB (%)**

Os respondentes do ensino fundamental regular apresentam maior conhecimento de tais situações para todos os grupos sociais pesquisados. Também, apresenta-se o IPCSB% por faixa etária.

Tabela A4.3 Conhecimento de Situações de Discriminação – IPCSB % - para Alunos de Ensino Fundamental

<b>Motivo</b>	<b>IPCSB%</b>
Negro	16% (Existe 16% de possibilidade de sofrer bullying por ser negro)
Pobre	15%
Homossexual	12%
Mulher	12%
Idoso	11%
Morador de periferia/favela	8%
Deficiente físico	7%
Deficiente mental	7%
Índio	5%
Cigano	4%

Tabela A4.4 Conhecimento de Situações de Discriminação – IPCSB % - para alunos até 14 anos

<b>Motivo</b>	<b>IPCSB%</b>
Negro	15,8% (Existe 15,8% de possibilidade de sofrer bullying por ser negro)
Pobre	15,6%
Homossexual	12%
Mulher	11,6%
Idoso	11,2%
Morador de periferia/favela	7,6%
Deficiente físico	7%
Deficiente mental	6,5%
Índio	4%
Cigano	3,6%

A literatura e experiências mostram que a mudança desse ambiente discriminatório marcadamente dissimulado leva muitos e muitos anos, possivelmente até gerações. No entanto, é preciso inicializar e potencializar esse processo por meios de ações corajosas, envolvendo disseminação de informações (condição necessária, mas não suficiente para a promoção de mudanças), realização de ações específicas e pontuais, implementação de plano, que visem à mudança de comportamento e, principalmente, no longo prazo, ações que promovam a mudança de valores dos agentes escolares em relação à questão discriminatória.

## A5. Modelo Estatístico do tipo Logit

### A5.1. Dados Obtidos inicialmente e variáveis do modelo

Dados inicialmente colhidos indicam os seguintes percentuais referentes aos grupos vítimas de “*bullying*” para alunos do Ensino Fundamental, conforme a Tabela A5.1:

Tabela A5.1 - Percentual das variáveis observadas nos grupos de vítimas.

<b>Grupo social vítima</b>	<b>Percentual (%)</b>
Negro	16
Pobre	15
Homossexual	12
Mulher	12
Idoso	11
Morador de periferia/favela	8
Deficiente físico	7
Deficiente mental	7 (*)
Índio	5
Cigano	4
Outros/não identificado	10

(\*) Neste texto não foi considerado esse ponto, mas isso não alterou a metodologia de análise. Esse estudo balizou a escolha das variáveis que têm o propósito de antecipar um possível potencial para o “*bullying*” e principalmente de seus pesos, que ainda podem (e devem) ser adequadamente calibrados, conforme a Tabela A5.2. Nessa tabela foram consideradas as seguintes variáveis: Raça/Grupos Sociais (V\_rac), Gênero (V\_gen), Classe Socioeconômica (V\_cla), Idade (V\_ida), e Deficiência Física (V\_def). Em cada variável foram colocados os grupos principais que foram detectados pela pesquisa inicial das situações de “*bullying*” e seus respectivos pesos (também relacionados às pesquisas das situações de “*bullying*”).



Tabela A5.2 - Variáveis do modelo.

Raça/Grupos Sociais - V_rac	Gênero - V_gen	Classe Socioec. V_cla	Idade - V_ida	Deficiência Física - V_def
Negro - 5	Transexual - 5	Favelado/Morador de Periferia - 6	> 60 anos - 5	Grave - 5
Mestiço/Mulato - 4	Homossexual - 4	Pobre Classe D - 5	40 a 60 anos - 4	Leve - 3
Índio - 3	Mulher - 3	Pobre Classe C - 4	20 a 40 anos - 2	Sem - 1
Cigano - 2	Homem - 1	Classe Média Baixa - 3	< 20 anos - 1	
Branco - 1		Classe Média - 2		
		Classe Média Alta/ Classe Alta - 1		

## A5.2. Simulação a partir dos dados iniciais

Embora as variáveis do modelo tenham sido obtidas a partir da análise de situações reais, os questionários não foram efetivamente aplicados. Dessa maneira, para efeito de simulação foram considerados os percentuais reais do trabalho da FIPE para os grupos correspondentes a cada variável, conforme a Tabela A5.3.

As variáveis tiveram pesos de 1 a 5 (com exceção de Classe Socioeconômica que pode chegar a 6), e os sorteios de 20 entrevistas foram feitos por meio do software Matlab, com código-exemplo, na Figura A5.1, relacionado à variável V\_rac. Dessa maneira, o sorteio de 20 “entrevistas” resultou na Tabela A5.4:

```

for i=1:20, a=rand*10
if a<30,V_rac(i)=5;
    else if a<50, V_rac(i)=4;
        else if a<65, V_rac(i)=3;
            else if a<70, V_rac(i)=2;
                else V_rac(i)=1; end,
            end,
        end,
    end,
end
end

```

Como não foram aplicados questionários reais, cabe ressaltar que a variável V\_resp foi obtida por meio de avaliação do número total dos pontos obtidos em cada uma das 20 entrevistas. Observando o histograma da Fig. 1, nota-se que a partir de 13 pontos o potencial de “bullying” se torna mais significativo. Dessa maneira, preenchemos a V\_resp da Tabela IV com o valor “1” quando o valor supera 13 pontos e “0”, caso contrário. Embora a simulação seja muito importante e ajude a entender

melhor o problema, o valor da V\_resp deverá ser obtido a partir da aplicação do questionário e/ou outras situações da prática.

Tabela A5.3 - Percentual das variáveis na ocorrência de “bullying”.

Raça/Grupos Sociais - V_rac	Gênero - V_gen	Classe Socioecon. V_cla	Idade - V_ida	Deficiência Física - V_def
Negro – 30%	Transexual - 4%	Favelado/Morador de Periferia – 10%	> 60 anos – 5%	Grave – 3%
Mestiço/ Mulato – 30%	Homossexual – 6%	Pobre Classe D – 20%	40 a 60 anos – 20%	Leve – 7%
Índio – 5%	Mulher – 45%	Pobre Classe C – 30%	20 a 40 anos – 30%	Sem – 90%
Cigano – 5%	Homem – 45%	Classe Média Baixa – 25%	< 20 anos – 45%	
Branco – 30%		Classe Média – 13%		
		Classe Média Alta/ Classe Alta – 2%		

Tabela A5.4 - “Entrevistas” obtidas por simulação.

“Entrevista”	V_rac	V_gen	V_cla	V_ida	V_def	V_resp
1	4	3	2	1	1	0
2	5	4	4	1	1	1
3	5	1	6	4	1	1
4	4	4	5	2	1	1
5	3	1	3	1	1	0
6	1	3	6	4	1	1
7	1	1	4	1	1	0
8	1	3	4	2	1	0
9	5	3	2	1	1	0
10	1	1	5	2	1	0
11	5	1	5	4	1	1
12	1	1	3	2	3	0
13	3	5	5	1	1	1
14	5	1	1	2	1	0
15	5	5	5	2	1	1
16	4	3	3	1	1	0
17	1	3	3	1	3	0
18	5	3	3	5	1	1
19	4	3	4	1	1	0
20	5	3	3	1	1	0

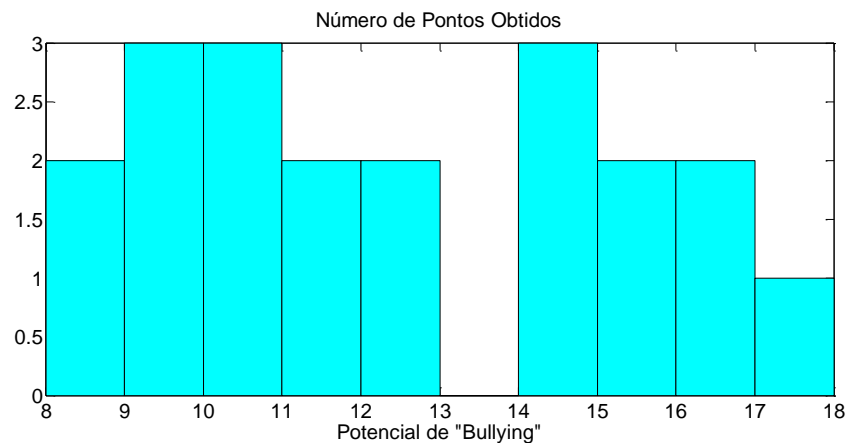


Figura A5.1 - Determinação do potencial de "bullying".

A seguir, o software Matlab foi utilizado para determinar as variáveis da regressão Logit, da seguinte maneira:

```
>> [b,d,s]=glmfit([V_gen V_rac V_cla V_ida V_def ], [V_resp Total],
'binomial'), fornecendo como resposta:
b = -705.1256  58.7698  39.3944  50.7115  55.7755  62.3866
d = 8.8818e-015
s = beta: [6x1 double] dfe: 14  sfit: 1.7810e-008  s: 1  estdisp: 0  covb: [6x6
double]  se: [6x1 double]  coeffcorr: [6x6 double]  t: [6x1 double]  p: [6x1 double]
resid: [20x1 double]  residp: [20x1 double]  residd: [20x1 double]  resida: [20x1
double]
```

Os resultados mostram um ajuste praticamente perfeito ( $d = 8.8818e-015$ ) à regressão Logit.

$$P(T_i \leq X_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5)}} =$$

$$= \frac{1}{1 + e^{-(705,1 + 58,8V_{gen} + 39,4V_{rac} + 50,7V_{cla} + 55,8V_{ida} + 62,4V_{def})}}$$

Calculamos EY a partir dos coeficientes do vetor **b** no Matlab como:

$$EY = 1 ./ (1 + \exp(-b(1) - b(2)*V_{gen} - b(3)*V_{rac} - b(4)*V_{cla} - b(5)*V_{ida} - b(6)*V_{def}))$$

Para confirmação dos resultados, a Tabela V coloca lado a lado o valor EY e a  $V_{resp}$ :

Tabela A5.5: Valores obtidos (EY) e V\_resp

EY	0	1	1	1	0	1	0	0	0	0	1	0	1	0	1	0	0	1	0	0
V_resp	0	1	1	1	0	1	0	0	0	0	1	0	1	0	1	0	0	1	0	0

### A5.3 Conclusões

A importância da aplicação deste modelo é principalmente devido à sua simplicidade, mas requer a utilização de um software do tipo *Matlab*. Podemos observar que o modelo deve ser calibrado com mais dados reais, mas ele já pode ser aplicado como um ajuste mais “grosso” em relação ao resultado final, mas que já pode ser um indicativo importante a partir de um questionário bastante simples. O modelo não prescinde de outros modelos, principalmente os baseados em “*Machine Learning*” e que podem ser utilizados para validação mútua e inclusive para a calibração de ambos (lembrando que os modelos baseados em Gerência do Conhecimento e/ou Inteligência Artificial dependem de uma etapa de aprendizado ou de treinamento).

## APÊNDICE B - RESULTADOS DA FASE DE TREINAMENTO DOS ALGORITMOS PARA A CRIAÇÃO DE NOVOS MODELOS

### B1. Pré-processamento

Os novos modelos foram criados no software WEKA<sup>13</sup> com os algoritmos de SVM disponíveis nos pacotes LIBLINEAR e LibSVM e o algoritmo de árvore de decisão J48. A *training set* está formado por 6,687 tweets rotulados (4,793 da classe “yes\_trace” *bullying* e 1,894 da classe “no\_trace” *bullying*). Na Tabela B1.1 se mostram os parâmetros do filtro “*StringToWordVector*” disponível no WEKA para o pré-processamento dos dados, principalmente este filtro permite transformar o texto em formato de vetor numérico.

Tabela B1.1 – Parâmetros do filtro “StringToWordVector”

Parâmetro	Valor
<i>IDF Transform</i>	<i>True</i>
<i>TF Transform</i>	<i>True</i>
<i>Attribute Indices</i>	<i>First-last</i>
<i>lowerCaseTokens</i>	<i>False</i>
<i>normalizeDocLength</i>	<i>Normalize All data</i>
<i>Stemmer</i>	<i>NullStemmer</i>
<i>Stopwords</i>	<i>Null</i>
<i>tokenizer</i>	<i>N-gram (unigram and bigram)</i>
<i>Words to Keep</i>	<i>4000</i>

Os resultados do treinamento desses algoritmos foram comparados e apresentados a seguir.

<sup>13</sup> <http://www.cs.waikato.ac.nz/ml/weka/>

## B2. Criação do modelo “*Bullying trace*”

O LibLINEAR apresentou um total de 86.4812% de instancias corretamente classificadas diante de 86.3167% (LibSVM) e 82.279%(J48), além disso o tempo para a criação do modelo foi de só 0,8 segundos e a execução do teste também foi rápido, tempo muito menor diante os outros dois classificadores.

### B2.1 Resultados da execução com LIBLINEAR

Tabela B1.2 – Parâmetros para SVM-LIBLINEAR

Parâmetros	Valor
<i>SVM Type</i>	<i>L2-regularized L2-loss support vector classification (dual)</i>
<i>Bias</i>	<i>1</i>
<i>Normalize</i>	<i>True</i>
<i>numDecimalPlaces</i>	<i>2</i>
<i>Probability Estimates</i>	<i>False</i>

```

Instances:      6687
Attributes:    7200
Test mode:     10-fold cross-validation
=== Classifier model (full training set) ===
Model bias=1.0 nr_class=2 nr_feature=7200 solverType=L2R_L2LOSS_SVC_DUAL
Time taken to build model: 0.8 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      5783           86.4812 %
Incorrectly Classified Instances    904           13.5188 %
Kappa statistic                     0.6624
Mean absolute error                  0.1352
Root mean squared error              0.3677
Relative absolute error              33.2927 %
Root relative squared error          81.603 %
Total Number of Instances          6687

=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precis  Recall  F-Meas  MCC  ROC Area  PRC Area  Class
0,915    0,261    0,899   0,915   0,907   0,663  0,827    0,883    yes_trace
0,739    0,085    0,774   0,739   0,756   0,663  0,827    0,646    no_trace
0,865    0,211    0,863   0,865   0,864   0,663  0,827    0,816

=== Confusion Matrix ===
  a    b  <-- classified as
4384  409 |    a = yes_trace
  495 1399 |    b = no_trace

```

## B2.2 Resultados da execução com LibSVM

Tabela B1.3 – Parâmetros para SVM-LibSVM

Parâmetros	Valor
Kernel Type	Linear
Normalize	True

```

Scheme:          weka.classifiers.functions.LibSVM -S 0 -K 0 -D 3 -G 0.0 -R 0.0
-N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -Z -model "C:\\Program Files\\Weka-3-
8" -seed 1
Instances:      6687
Attributes:     7200
Test mode:     10-fold cross-validation

```

=== Classifier model (full training set) ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

Time taken to build model: **29.89 seconds**

=== Stratified cross-validation ===

=== Summary ===

```

Correctly Classified Instances      5772           86.3167 %
Incorrectly Classified Instances    915           13.6833 %
Kappa statistic                    0.6594
Mean absolute error                 0.1368
Root mean squared error             0.3699
Relative absolute error             33.6978 %
Root relative squared error        82.098 %
Total Number of Instances          6687

```

=== Detailed Accuracy By Class ===

```

TP Rate  FP Rate  Precis  Recall  F-Meas  MCC    ROC Area  PRC Area  Class
0,912    0,259    0,899   0,912   0,905   0,660  0,826    0,883    yes_trace
0,741    0,088    0,768   0,741   0,754   0,660  0,826    0,642    no_trace
0,863    0,211    0,862   0,863   0,862   0,660  0,826    0,815

```

=== Confusion Matrix ===

```

  a    b  <-- classified as
4369  424 |    a = yes_trace
 491 1403 |    b = no_trace

```

## B2.3 Resultados da execução com J48

```

Instances:      6687
Attributes:     7200
Test mode:     10-fold cross-validation
Number of Leaves :    376
Size of the tree :    751

```

Time taken to build model: **475.07 seconds**

=== Summary ===

Correctly Classified Instances	5502	<b>82.279%</b>
Incorrectly Classified Instances	1185	17.721%
Kappa statistic	0.5565	
Mean absolute error	0.2031	
Root mean squared error	0.397	
Relative absolute error	50.0121 %	
Root relative squared error	88.1162 %	
Total Number of Instances	6687	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precis	Recall	F-Meas	MCC	ROC Area	PRC Area	Class
0,887	0,339	0,869	0,887	0,878	0,557	0,795	0,856	yes_trace
0,661	0,113	0,697	0,661	0,679	0,557	0,795	0,618	no_trace
0,823	0,275	0,820	0,823	0,821	0,557	0,795	0,788	

=== Confusion Matrix ===

```

a      b      <-- classified as
4250  543 |      a = yes_trace
642   1252 |      b = no_trace

```

### B3. Criação do modelo “*Bullying Form*”

Os três classificadores obtiveram mais do 90% das instâncias corretamente classificadas, sendo o algoritmo J48 quem teve a melhor porcentagem (96.2824%) mas demorou 92.15seg para a criação do modelo e muito mais tempo para a execução dos testes. Contudo, não pode-se afirmar a efetividade dos classificadores porque o *training set* para o caso de tipo de *bullying*, não estava bem balanceado devido a que tinham mais instâncias da classe “*general*”. Analisando as outras medidas como o *f-measure*, observa-se que para as classes “*physical*”, “*verbal*” e “*cyberbullying*”, está baixo de 60%. Resultados similares foram apresentados na pesquisa de (XU, 2015).

#### B3.1 Resultados da execução com LIBLINEAR

```

Instances:      4788
Attributes:     7487
Time taken to build model: 0.77 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      4506      94.1103 %
Incorrectly Classified Instances     282      5.8897 %
Kappa statistic                     0.464
Mean absolute error                  0.0294
Root mean squared error              0.1716
Relative absolute error              37.975 %

```



```

Root relative squared error          87.3046 %
Total Number of Instances           4788

```

```
=== Detailed Accuracy By Class ===
```

TP Rate	FP Rate	Precis	Recall	F-Meas	MCC	ROC Area	PRC Area	Class
0,997	0,645	0,945	0,997	0,971	0,552	0,676	0,945	general
0,273	0,004	0,746	0,273	<b>0,400</b>	0,439	0,635	0,233	physical
0,431	0,001	0,917	0,431	<b>0,587</b>	0,624	0,715	0,408	cyberbul
0,260	0,001	0,781	0,260	<b>0,391</b>	0,445	0,629	0,218	verbal
0,941	0,593	0,933	0,941	0,928	0,547	0,674	0,890	

```
=== Confusion Matrix ===
```

	a	b	c	d	<-- classified as
4384	7	2	3	3	a = general
136	53	2	3	3	b = physical
53	4	44	1	1	c = cyberbullying
64	7	0	25	1	d = verbal

## B3.2 Resultados da execução com LIBSVM

```

Instances:      4788
Time taken to build model: 3.78 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      4480          93.5673 %
Incorrectly Classified Instances    308          6.4327 %
Kappa statistic                    0.3843
Mean absolute error                 0.0322
Root mean squared error             0.1793
Relative absolute error             41.4762 %
Root relative squared error         91.2405 %
Total Number of Instances          4788

```

```
=== Detailed Accuracy By Class ===
```

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,997	0,727	0,939	0,997	0,967	0,469	0,635	0,939	general
0,258	0,003	0,781	0,258	0,388	0,437	0,627	0,231	physic
0,314	0,001	0,889	0,314	0,464	0,523	0,656	0,293	cyberbu
0,177	0,001	0,773	0,177	0,288	0,365	0,588	0,153	verbal
0,936	0,668	0,928	0,936	0,919	0,467	0,634	0,881	

```
=== Confusion Matrix ===
```

	a	b	c	d	<-- classified as
4381	9	3	3	3	a = general
142	50	1	1	1	b = physical
68	1	32	1	1	c = cyberbullying
75	4	0	17	1	d = verbal

## B3.3 Resultados da execução com J48

```

Instances:      4788
Number of Leaves :      44
Size of the tree :      87

```

Time taken to build model: **92.15 seconds**

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	4610	<b>96.2824 %</b>
Incorrectly Classified Instances	178	3.7176 %
Kappa statistic	0.7464	
Mean absolute error	0.0241	
Root mean squared error	0.131	
Relative absolute error	31.0606 %	
Root relative squared error	66.6457 %	
Total Number of Instances	4788	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,994	0,173	0,985	0,994	0,989	0,864	0,951	0,991	general
0,701	0,015	0,657	0,701	0,678	0,665	0,879	0,493	physical
0,578	0,003	0,787	0,578	0,667	0,669	0,840	0,573	cyberbul
0,469	0,005	0,662	0,469	0,549	0,549	0,795	0,420	verbal
0,963	0,160	0,961	0,963	0,961	0,846	0,942	0,951	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
4370	22	1	3	a = general
29	136	11	18	b = physical
27	14	59	2	c = cyberbullying
12	35	4	45	d = verbal

## B4. Criação do modelo “Author Role”

### B4.1 Resultados da execução com LIBLINEAR

Instances: 4788

Attributes: 7894

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Model bias=1.0 nr\_class=5 nr\_feature=7894 solverType=L2R\_L2LOSS\_SVC\_DUAL

Time taken to build model: 1.5 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	4312	<b>90.0585 %</b>
Incorrectly Classified Instances	476	9.9415 %
Kappa statistic	0.8042	
Mean absolute error	0.0398	
Root mean squared error	0.1994	
Relative absolute error	19.0547 %	
Root relative squared error	61.7519 %	
Total Number of Instances	4788	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,283	0,001	0,875	<b>0,283</b>	0,427	0,492	0,641	0,262	bully
0,945	0,105	0,922	<b>0,945</b>	0,933	0,843	0,920	0,903	victim
0,030	0,001	0,400	0,030	0,056	0,106	0,515	0,026	accuser
0,000	0,000	0,000	0,000	0,000	-0,001	0,500	0,003	defender
0,908	0,086	0,872	<b>0,908</b>	0,890	0,816	0,911	0,828	reporter

0,901 0,094 0,891 0,901 0,890 0,812 0,903 0,845

=== Confusion Matrix ===

```

a   b   c   d   e   <-- classified as
28  37   1   0  33 |   a = bully
 4 2570   0   0 147 |   b = victim
 0   8   2   0  57 |   c = accuser
 0   1   1   0  14 |   d = defender
 0 171   1   1 1712 |   e = reporter

```

## B4.2 Resultados da execução com LIBSVM

Instances: 4788  
 Attributes: 7894  
 Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

Time taken to build model: 7.95 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	4299	<b>89.787%</b>
Incorrectly Classified Instances	489	10.213%
Kappa statistic	0.7988	
Mean absolute error	0.0409	
Root mean squared error	0.2021	
Relative absolute error	19.5751 %	
Root relative squared error	62.5895 %	
Total Number of Instances	4788	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,283	0,001	0,848	0,283	0,424	0,485	0,641	0,255	bully
0,943	0,107	0,921	0,943	0,932	0,840	0,918	0,901	victim
0,015	0,001	0,250	0,015	0,028	0,058	0,507	0,018	accuser
0,000	0,000	0,000	0,000	0,000	-0,001	0,500	0,003	defender
0,903	0,089	0,868	0,903	0,885	0,809	0,907	0,822	reporter
0,898	0,096	0,886	0,898	0,887	0,807	0,901	0,841	

=== Confusion Matrix ===

```

a   b   c   d   e   <-- classified as
28  36   0   0  35 |   a = bully
 5 2567   0   0 149 |   b = victim
 0   5   1   0  61 |   c = accuser
 0   2   0   0  14 |   d = defender
 0 178   3   1 1703 |   e = reporter

```

## B4.3 Resultados da execução com J48

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Instances: 4788

Attributes: 7894

Time taken to build model: 397.11 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	4175	<b>87.1972 %</b>
--------------------------------	------	------------------

---

Incorrectly Classified Instances	613	12.8028 %
Kappa statistic	0.752	
Mean absolute error	0.0632	
Root mean squared error	0.2164	
Relative absolute error	30.2694 %	
Root relative squared error	67.0231 %	
Total Number of Instances	4788	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,596	0,006	0,663	0,596	0,628	0,621	0,846	0,508	bully
0,917	0,157	0,885	0,917	0,900	0,764	0,892	0,872	victim
0,657	0,004	0,677	0,657	0,667	0,662	0,815	0,449	accuser
0,188	0,001	0,429	0,188	0,261	0,282	0,793	0,104	defender
0,836	0,080	0,871	0,836	0,853	0,761	0,892	0,817	reporter
0,872	0,121	0,870	0,872	0,871	0,757	0,890	0,834	

=== Confusion Matrix ===

a	b	c	d	e	<-- classified as
59	20	11	1	8	a = bully
12	2494	1	0	214	b = victim
14	1	44	2	6	c = accuser
0	0	8	3	5	d = defender
4	304	1	1	1575	e = reporter

---

# APÊNDICE C – ARTIGO: APLICAÇÃO DO *MIT APP INVENTOR* COMO FERRAMENTA DE APOIO À APRENDIZAGEM

---

O artigo “Aplicação do *MIT App Inventor* como ferramenta de apoio à aprendizagem”, foi apresentado no CIEDUC 2017<sup>14</sup> - “IX Congreso Iberoamericano de Educación Científica” na cidade de Mendoza, Argentina o dia 15 de março de 2017. Esse artigo foi parte do trabalho realizado na disciplina de Gestão de Projetos da Faculdade de Tecnologia (FT) durante o semestre 2016-2, com supervisão do Prof. Dr. Marco Borges.

## Aplicação do *MIT App Inventor* como ferramenta de apoio à aprendizagem

Pamela Rosales Sedano<sup>1</sup>, Igor Vasconcelos Nogueira<sup>1</sup>, Sheila Vilas Boas Fraga<sup>1</sup>, Marcos Borges<sup>1</sup>

<sup>1</sup>Faculdade de Tecnologia

Universidade Estadual de Campinas (UNICAMP)

Limeira (São Paulo) - Brasil

c154133@g.unicamp.br, igor@ifspcapivari.com.br, sheilavbfbio@gmail.com, [marcos@ft.unicamp.br](mailto:marcos@ft.unicamp.br)

**Resumo.** O conhecimento é um componente imprescindível ao desenvolvimento humano, econômico e social. Associado à tecnologia, o conhecimento é fundamental para o desenvolvimento das nações. Neste contexto, a inserção do Pensamento Computacional no ambiente escolar como processo evolutivo do ensino e aprendizagem é uma permanente necessidade brasileira. Com esse foco, foi desenvolvido o projeto de criação de aplicativos por meio da ferramenta *MIT App Inventor*, em uma turma de jovens entre 14 e 17 anos. Como ferramenta de avaliação do projeto foram aplicados dois questionários e foi possível verificar que 60% dos alunos tinham a percepção inicial que programar computadores era difícil, 25% achavam muito difícil, 5% se mostravam indiferentes e apenas 10% consideravam fácil. Após os jovens criarem os aplicativos no *MIT App Inventor*, foi verificado que 80% estavam interessados ou muito interessados em aprender mais sobre o *MIT App Inventor*, 10% se mostraram indiferentes e outros 10% indicaram ter pouco interesse. Com relação ao Pensamento Computacional os resultados mostraram que aproximadamente 69% concordam em continuar empregando a metodologia do Pensamento Computacional na resolução de problemas em outras atividades, 25% não sabem se irão adotar esta metodologia e 6,3% discordam parcialmente no uso desta metodologia para outras atividades.

**Palavras chaves:** MIT App Inventor; Pensamento Computacional; TIC

### 1. Introdução

---

<sup>14</sup> CIEDUC 2017 - <http://www.cieduc.org/2017/>

O papel da informação e da comunicação no processo de transformação política, econômica e social vivido pela sociedade se destaca no momento atual, assim como as tecnologias que estão cada vez mais presentes no cotidiano da sociedade [1].

Nesse sentido, torna-se relevante a proposta de introduzir nas matrizes curriculares dos programas escolares o uso da metodologia do do Pensamento Computacional com o auxílio das Tecnologias de Informação e Comunicação (TIC). Tal metodologia visa contribuir para o desenvolvimento das habilidades cognitivas, sociais e profissionais exigidas na atualidade e estão em acordo com o que postula o documento Sociedade da Informação no Brasil [2] quando evidencia a necessidade de se adotar metodologias mais adequadas para a integração das TIC na escola [1].

Segundo [3] “[...] nossos alunos mudaram radicalmente. Os alunos de hoje não são os mesmos para os quais o nosso sistema educacional foi criado [...]”. Os autores de [4] corroboram com esta visão ao afirmarem que em acordo com tais tendências, as tecnologias necessitam ser integradas, inclusive, nos processos de ensino-aprendizagem, remodelando e redesenhando novos hábitos e métodos presentes em tais processos.

Apesar de existir muitos estudos sobre o auxílio na melhoria do ensino-aprendizagem utilizando ferramentas tecnológicas (computadores e outros aparelhos eletrônicos) e a Internet, faz-se necessário propor estratégias metodológicas e pedagógicas a fim de integrar de maneira eficaz as tecnologias nos processos educacionais. Uma estratégia metodológica e pedagógica que pode ser adotada para esse propósito é o “Pensamento Computacional” (CT – *Computational Thinking*), ainda incipiente e pouco promovido nas escolas de ensino fundamental e médio. Raciocinar logicamente para resolver problemas é uma habilidade desenvolvida que possibilita recursos cognitivos importantes para várias áreas do conhecimento [5].

O CT não tem como objetivo tornarem as crianças programadores, mas desenvolver nelas habilidades para resolução de problemas, *design* de sistemas e entendimento do comportamento humano através de conceitos fundamentais das Ciências da Computação (CS) [6]. A habilidade de pensar de forma computacional é essencial para qualquer disciplina, sendo importante expor as noções relacionadas ao CT aos alunos de escolas [7].

Dado a facilidade dos jovens para interagirem com computadores e *smartphones*, este artigo tem por **objetivo avaliar o uso do MIT App Inventor como ferramenta de apoio à aprendizagem, auxiliando no ensino dos alunos.**

A justificativa deste trabalho se dá pela incipiência de projetos que possibilitam promover o CT nos jovens na faixa etária de 14 à 17 anos. Assim, o uso do *MIT App Inventor* como ferramenta para o auxílio na aprendizagem, através do desenvolvimento do CT, permitindo desenvolver a capacidade enfrentar e resolver qualquer tipo de problema sendo útil nas diferentes disciplinas da educação formal. Além disso, reconhecer o potencial transformador que o CT pode trazer ao processo de ensino-aprendizagem a possibilidade de reorganizar as matrizes curriculares dos programas escolares de forma que estes apoiem o desenvolvimento das habilidades cognitivas, sociais e profissionais desejadas [8].

Após essa seção introdutória o trabalho foi estruturado em mais cinco seções. A seção 2 apresentou a revisão bibliográfica, enquanto a seção 3 os materiais e métodos. A seção 4 abordou as dinâmicas e aplicativos desenvolvidos e a seção 5 apresentou os resultados obtidos. Por fim, a seção 6 relatou as considerações finais.

## 2. Revisão Bibliográfica

A computação possibilita continuamente o avanço da inovação e pesquisa em todas as áreas, pois facilita os esforços para resolver problemas nas mais diversas áreas, por exemplo, na prevenção ou cura de doenças graves, além de expandir a compreensão de nós mesmos como sistemas biológicos e a relação com o mundo que nos rodeia [9].

De acordo com [10] [7], o CT é uma habilidade fundamental do pensamento analítico para todas as pessoas de diferentes faixas etárias, pois ajuda a resolver problemas com diferentes soluções, projetar sistemas e compreender o comportamento humano baseando-se nos conceitos fundamentais da Ciências da Computação.

Em [7] destacou-se que um relatório sobre o CT, emitido pelo Conselho Nacional de Pesquisa (NRC) constatou que os alunos podem aprender estratégias do CT enquanto estudam uma disciplina de qualquer área e que a orientação adequada dos professores pode permitir que os alunos aprendam a usar essas estratégias de forma independente.

Um ponto de relevância é a formação de professores. Estes precisam de recursos que demonstrem como integrar adequadamente esses novos conceitos do CT, primeiro em sua própria esfera de conteúdo pedagógico, e depois na prática em sala de aula [9].

Segundo [11][12] é importante inserir ferramentas que forneçam o suporte necessário para a aprendizagem de lógica de programação aos alunos de ensino fundamental e médio, com atividades lúdicas, a exemplo de jogos digitais. Em [4] utilizaram o *MIT App Inventor*, que propicia um ambiente de aprendizagem baseado no construcionismo, haja vista que permite aos alunos criarem aplicações à medida que descobrem e exercitam sua criatividade, tornando o aprendizado mais lúdico.

Diante das dificuldades no processo de ensino-aprendizagem do CT, em [13] utilizaram a abordagem *'blended learning'*, que utiliza ferramentas tecnológicas associadas às abordagens pedagógicas, tais como construcionismo e construtivismo. O trabalho em [13] utilizou o *MIT App Inventor* e verificou que o uso da linguagem de programação visual foi um facilitador no processo de aprendizagem, já que os estudantes não precisavam aprender a sintaxe da linguagem de programação tradicional.

Embora haja um amplo reconhecimento de que a computação permeia todos os aspectos da economia global, seu lugar como parte obrigatória do currículo escolar está longe de ser consolidado [6].

### 3. Materiais e Métodos

O *MIT App Inventor* é um *software* de introdução à programação e criação de aplicativos móveis. A vantagem de utilizar este tipo de *software* é motivar aos jovens a não serem simples consumidores de tecnologia senão tornando-os desenvolvedores. A programação utilizada no *MIT App Inventor* é através da funcionalidade de arrastar e soltar blocos ao invés da linguagem de código baseado em texto. O *MIT App Inventor* é muitas vezes chamado de programação baseada em **eventos** o que significa que o aplicativo funciona com base em reações a eventos entre o usuário e o **aplicativo e seus componentes** (por exemplo, clicar o componente botão, deslizar o dedo na tela do smartphone, entre outros). Portanto, a interface gráfica do *MIT App Inventor* permite aos usuários iniciantes criar aplicações básicas e cheias de funcionalidades em pouco tempo.

Os quatro conceitos do CT utilizados durante a criação dos aplicativos por meio do *MIT App Inventor* foram:

- Capacidade de decomposição de problemas: decomposição dos processos em peças menores para serem manipuladas com facilidade;
- Reconhecimento de padrões: observação de padrões, tendências e regularidades nos dados;
- Abstração: identificação dos princípios gerais que geram os padrões;
- Design* de algoritmos: desenvolvimento passo por passo de instruções para resolver problemas.

Assim, a metodologia adotada apoia-se nos pressupostos relacionados à pedagogia de projetos e construtivista, que possibilitam o “aprender-fazendo” e o reconhecimento da própria autoria no que foi produzido, por meio de questões que incentivam a contextualização dos conceitos aprendidos e a descoberta de outros conceitos que surgem durante o desenvolvimento do projeto [8].

### 4. Dinâmicas e Aplicativos Desenvolvidos

O presente trabalho consistiu na criação de aplicativos por meio do *software MIT App Inventor* e do desenvolvimento do CT em uma turma de 20 jovens entre 14-17 do Centro de Aprendizado Metódico e Prático de Limeira (CAMPL – Patrulheiro) que é uma instituição não governamental com objetivo de educar os adolescentes que se encontram em situação de vulnerabilidade socioeconômica.

A aplicação prática do projeto foi realizada por cinco encontros presenciais. O primeiro encontro com os alunos foi uma apresentação geral do projeto e do *MIT App Inventor* assim como motivá-los a programarem por meio de o vídeo “É muito difícil aprender a programar computadores?” criado pela Code.org®<sup>15</sup>. O roteiro dos seguintes encontros foi dividido da seguinte forma:

**Dinâmica 1:** Apresentação dos quatro conceitos da metodologia do CT e exemplos práticos sem utilizar o computador. Logo, apresentação da interface do *MIT App Inventor* por meio do desenvolvimento do aplicativo “Coelho na Cartola”<sup>16</sup>.

**Dinâmica 2:** Desenvolvimento dos aplicativos “Quem é esse *Pokémon*?”, “Calculadora Simples”<sup>17</sup> e “Somar de 1 até N”.

**Dinâmica 3:** Desenvolvimento do “Calculadora de Massa Corporal”<sup>18</sup>

**Dinâmica 4:** Desenvolvimento dos aplicativos “Bola Mágica 8”<sup>19</sup>, “Digital Doodle”<sup>20</sup>, “Vamos tirar uma foto”<sup>21</sup> focados nos eventos próprios de um *smartphone*.

Na metodologia explicada na seção anterior, ressalta-se o aprendiz como um participante ativo no processo de aprendizagem. Nessa premissa, escolheram-se aplicativos a serem desenvolvidos os quais ajudassem aos alunos a experimentar e descobrir seus próprios erros e acertos. Os aplicativos escolhidos (tabela 1) ensinaram aos alunos de forma gradual os conceitos iniciais de programação como: variáveis, estruturas de controle, estruturas de

<sup>15</sup> <https://code.org/>

<sup>16</sup> MagicTrick - <http://explore.appinventor.mit.edu/sites/all/files/teachingappcreation/unit1/MagicTrick.pdf>

<sup>17</sup> Aplicativo desenvolvido pelo grupo LIAG da FT-UNICAMP

<sup>18</sup> Aplicativo desenvolvido pelo grupo LIAG da FT-UNICAMP

<sup>19</sup> Magic 8-ball <http://appinventor.mit.edu/explore/ai2/magic-8-ball.html>

<sup>20</sup> Digital Doodle - <http://appinventor.mit.edu/explore/sites/all/files/hourofcode/DigitalDoodle.pdf>

<sup>21</sup> Paint Pot - <http://www.appinventor.org/apps2/paintpot2/paintpot2.pdf>

decisão e listas. Pesquisamos aplicativos que sejam de interesse dos alunos e relacionados a suas experiências atuais, todos eles sem fins comerciais.

O método ensinado aos alunos para o desenvolvimento de um aplicativo consistiu em (1) Explicar qual é o objetivo do aplicativo; (2) aplicação dos quatro conceitos do CT para resolver um problema e conseguir atingir o objetivo previamente mencionado; ressaltou-se aos alunos que o último conceito, definição do algoritmo, é a base para já começar a programar no *MIT App Inventor* ou qualquer linguagem de programação; (3) Apresentação da tela a ser criada pelos alunos e identificação dos componentes dela como botões, caixas de textos, legendas, etc. a serem utilizados; (4) criação da tela do aplicativo na interface “*Designer*” do *MIT APP Inventor* por meio do arrasto de componentes e ajuste de suas propriedades; (5) montagem dos blocos na interface “*Blocos*” do *MIT App Inventor* utilizando o algoritmo já definido; (6) Teste com o simulador, código QR ou executando no smartphone; (7) Desafio aos alunos que consistiu em melhoras ao aplicativo para avaliar seu entendimento e sua criatividade.

**Tabela 1.** Aplicativos desenvolvidos no *MIT App Inventor* pelos alunos do CAMPL

APLICATIVO	DESCRIÇÃO	META
<b>Coelho na Cartola</b>	Consiste em fazer um coelho aparecer na cartola quando clicar nela.	Conhecer as interfaces gráficas dos modos “ <i>Designer</i> ” e “ <i>Blocos</i> ” do <i>MIT App Inventor</i> .
<b>Quem é esse pokémon?</b>	Jogo de perguntas e respostas desenvolvido pela equipe deste projeto.	Praticar com os componentes “botão”, “legenda” e “multimídia”. Praticar com os blocos de controle e eventos
<b>Calculadora Simples</b>	Permitir somar, subtrair, multiplicar e dividir dois números.	Conhecer o componente “caixa de texto” e o bloco de matemáticas.
<b>Somar de 1 até N</b>	Calcular a soma de 1 até um número “N” ingressado pelo usuário.	Uso do bloco de matemáticas. Conhecer os blocos de decisão e variáveis.
<b>Calculadora de IMC</b>	Calcular o índice de massa corporal de acordo ao peso e altura.	Uso do bloco de matemáticas e variáveis. Praticar com os blocos de decisão.
<b>Bola mágica 8</b>	Mostrar e reproduzir por fala uma mensagem aleatória gerada pelo usuário por meio da sacudida do smartphone.	Conhecer os componentes “sensor acelerômetro” e “texto para falar” e o bloco de “listas”
<b>Digital Doodle</b>	Desenhar linhas por meio dos dedos na tela do smartphone.	Conhecer o componente “pintura” (canvas).
<b>Vamos tirar uma foto</b>	Tirar fotos para fazer desenhos sobre elas.	Conhecer o componente “câmera” e o bloco de “cores”

## 5. Resultados

A avaliação do projeto foi feita por meio de dois questionários criados no *Google Forms*. O primeiro foi aplicado no final da terceira dinâmica e constou de 18 perguntas, este questionário foi respondido pelos 20 alunos participantes (10 homens e 10 mulheres) e avaliou a apresentação geral, o entendimento dos quatro conceitos do CT e os cinco aplicativos desenvolvidos até esse momento. O segundo questionário foi aplicado no final da quarta e última dinâmica e constou de 13 perguntas, o qual foi respondido por 16 (6 homens e 10 mulheres) dos 17 alunos participantes desse dia e avaliou os três aplicativos desenvolvidos assim como a metodologia do CT.

### 5.1 Avaliação da apresentação geral

Na apresentação geral do projeto foi exibido o vídeo “É muito difícil aprender a programar computadores?” criado pela Code.org®. Avaliou-se como reagiram os alunos após de assistir o vídeo por meio da pergunta, “No primeiro dia foi exibido o vídeo “É muito difícil aprender a programar computadores?”, como parte do aprendizado do *MIT App Inventor*. Como você se sentiu logo após assistir o vídeo?”. Os resultados foram: 65% dos alunos respondeu que se sentiu motivado e 35% muito motivado para aprender a programar, o qual era a meta da equipe com o vídeo.

Avaliou-se, a percepção inicial dos alunos sobre a programação por meio da pergunta “Antes de ver o vídeo, o que você achava sobre aprender a programar computadores?”, porque a maioria das pessoas percebem a programação de computadores como uma atividade muito técnica e apropriada só para um pequeno segmento da



população ou que a sintaxe das linguagens de programação é muito difícil de aprender [14]. Os resultados confirmaram essa percepção na turma já que 60% achou que programar computadores é difícil, 25% achou muito difícil, enquanto 5% se mostrou indiferente e só 10% achou fácil.

## 5.2 Avaliação do MIT App Inventor e os aplicativos desenvolvidos

A primeira experiência dos alunos com o MIT App Inventor foi avaliada por meio da pergunta “Nosso primeiro aplicativo *“Coelho na Cartola”* foi utilizado para conhecer a interface do MIT App Inventor. Como foi essa primeira experiência?”. As 5 possíveis respostas para esta questão na escala de Likert foram “Muito difícil”, “difícil”, “nem fácil nem difícil”, “fácil” e “muito fácil” e os resultados mostraram que 50% achou fácil ou muito fácil o uso do MIT App Inventor, 35% achou nem fácil nem difícil e só 15% achou difícil.

A avaliação do nível dos aplicativos foi por meio da questão: “Para você, qual foi o nível deste aplicativo?” com as mesmas respostas da questão anterior. Os resultados para o aplicativo “**Coelho na Cartola**” mostraram que 70% dos alunos acharam que o nível foi fácil ou muito fácil, 25% o acharam nem fácil nem difícil, e só 5% (um dos 20 alunos) achou que foi muito difícil. No caso do aplicativo “**Quem é esse pokémon**”, 45% dos alunos acharam que o nível foi fácil ou muito fácil, 35% o acharam nem fácil nem difícil, e 20% achou que foi difícil ou muito difícil. Os resultados para o aplicativo “**Somar de 1 até N**” mostraram que 45% dos alunos acharam que o nível foi fácil ou muito fácil, 30% o acharam nem fácil nem difícil, e 25% achou que foi difícil.

Os resultados para o aplicativo “**Bola Mágica 8**” mostraram que 81,3% dos alunos acharam que o nível foi fácil ou muito fácil, 12,5% acharam nem fácil nem difícil, e 6,3% achou que foi difícil. O aplicativo “**Vamos tirar foto**” teve como resultados que 68,8% dos alunos acharam que o nível foi fácil ou muito fácil, 25% acharam nem fácil nem difícil, e 6,3% achou que foi difícil.

O desempenho dos alunos durante o desenvolvimento dos aplicativos foi avaliado por meio da questão: “Você conseguiu acompanhar o aplicativo?” com 5 possíveis respostas na escala Likert: “Não consegui acompanhar e não terminei”, “Consegui acompanhar com dificuldade e terminei mas não entendi o que eu estava fazendo”, “Consegui acompanhar, terminei e entendi o que eu estava fazendo”, “Foi fácil, consegui acompanhar, terminei rápido e entendi o que eu estava fazendo” e “Foi muito fácil, consegui acompanhar, terminei bem rápido e entendi o que eu estava fazendo”.

Os resultados para o aplicativo “**Coelho na Cartola**” mostraram que 100% dos alunos conseguiu acompanhar e terminar o aplicativo, 45% do total considerou fácil (25%) ou muito fácil (20%) o desenvolvimento do aplicativo, embora teve 10% dos alunos que terminou o aplicativo mas não conseguiu entender o que estava fazendo. Os resultados do aplicativo “**Quem é esse Pokémon**” mostraram que 80% dos alunos conseguiu acompanhar e terminar o aplicativo entendendo o que fazia, 15% conseguiu acompanhar mas não entendeu o que estava fazendo e 5% não conseguiu acompanhar o aplicativo. No caso do aplicativo “**Somar 1 até N**” os resultados mostraram que 85% dos alunos conseguiu acompanhar e terminar o aplicativo entendendo o que estava fazendo, 15% conseguiu acompanhar mas não entendeu o que estava fazendo (fig. 1).

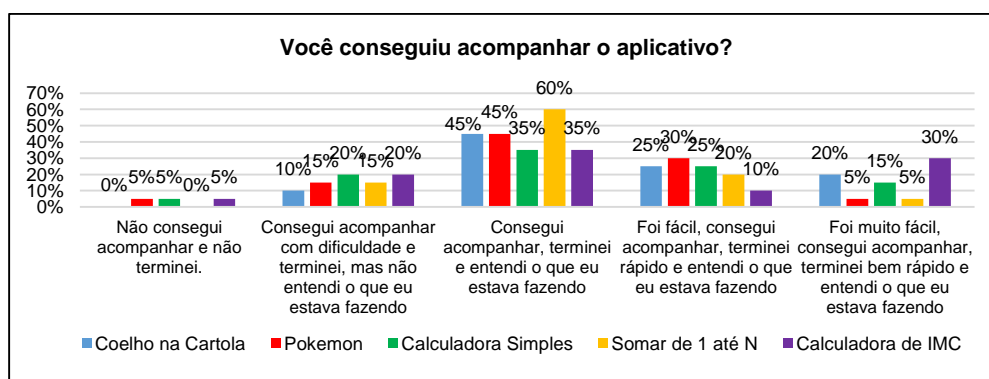


Fig. 1. Avaliação do desempenho no desenvolvimento dos cinco primeiros aplicativos

Os resultados para o aplicativo “**Bola mágica 8**” mostraram que 94% dos alunos conseguiu acompanhar e terminar o aplicativo entendendo o que estava fazendo e um 6% não conseguiu terminar. Os resultados para o aplicativo “**Vamos tirar uma foto**” mostraram que 86,6% dos alunos conseguiu acompanhar e terminar o aplicativo entendendo o que estava fazendo e um 13,3% conseguiu acompanhar mas não entendeu o que estava fazendo (fig. 2).

O interesse em continuar programando no MIT App Inventor foi avaliado por meio da questão “Após criar os aplicativos, qual o seu interesse de aprender mais sobre o MIT App Inventor?”. O 80% dos alunos estava interessado ou muito interessado em aprender mais sobre o MIT App Inventor, 10% se mostrou indiferente e 10% mostrou ter pouco interesse.

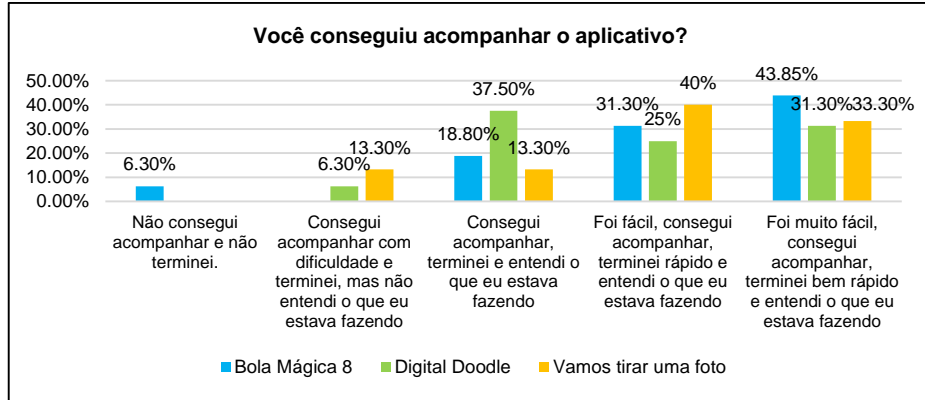


Fig. 2. Avaliação do desempenho no desenvolvimento dos aplicativos na quarta dinâmica

5.3 Avaliação da metodologia do Pensamento Computacional (CT)

Os autores orientaram aos alunos para desenvolverem os aplicativos utilizando os conceitos do CT como uma boa prática para resolução de problemas e uma habilidade fundamental do pensamento analítico para qualquer um, não só para aqueles focados na CS [7].

Dessa forma, avaliou-se o entendimento dos alunos sobre os conceitos do CT e a aplicação destes para resolver um problema por meio da questão (no questionário 1) “Na primeira dinâmica foram apresentados os quatro conceitos necessários para resolver um problema: decomposição; reconhecimento de padrões; abstração e algoritmos. Para você, eles o ajudaram a obter uma solução”. Os resultados do questionário mostraram que 70% dos alunos entenderam os conceitos e como eles podem resolver os problemas, 20% entendeu os conceitos mas não conseguiu entender como eles podem resolver um problema e 10% só entendeu alguns dos quatro conceitos portanto não conseguiu entender como aplicá-los.

A metodologia do CT utilizada neste projeto foi avaliada por meio das seguintes três questões apresentadas (fig.3), com cinco possíveis respostas na escala de Likert<sup>22</sup> : “Quanto você gostaria de ter mais aulas que trabalhasse essa metodologia do CT?”, “o quanto você acredita que essa metodologia do CT poderia ajudá-lo a resolver problemas em outras disciplinas do seu curso?” e “o quanto você acredita que continuará adotando essa metodologia do CT para ajudá-lo a resolver problemas em outras atividades?”.

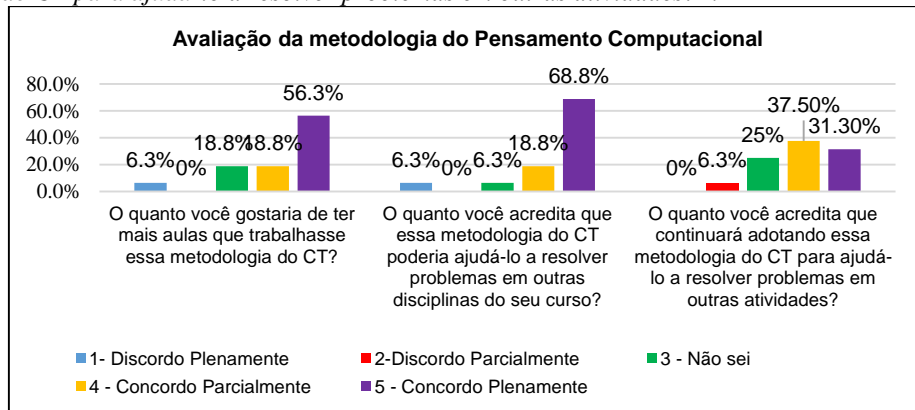


Fig. 3. Avaliação da metodologia do Pensamento Computacional

Na primeira questão, os resultados mostraram que 75,1% concorda parcial ou plenamente de que gostariam ter mais aulas com a metodologia do CT, 18,8% não sabe e só um 6,3% não gostaria ter mais aulas com essa metodologia. Os resultados para a segunda questão mostraram que 87,6% concorda que é possível utilizar a metodologia do CT em outras disciplinas, não só aquela relacionada à computação, no curso, 6,3% não sabe e o outro 6,3% não concorda com a utilidade desta metodologia em outras disciplinas.

Finalmente, na terceira questão, os resultados mostraram que 68,8% concorda parcial ou plenamente que continuará adotando esta metodologia para resolver problemas em outras atividades, 25% não sabe e 6,3% discorda parcialmente no uso desta metodologia para outras atividades.

<sup>22</sup> 1-“Discordo Plenamente”, 2-“Discordo Parcialmente”, 3- “Não sei”, 4 –“Concordo Parcialmente” e 5-“Concordo Plenamente”

## 6. Considerações Finais

A apresentação do vídeo motivacional feito pela Code.org ajudou muito aos alunos mudarem a ideia “mitificada” da programação e despertar neles o interesse para começar a programar no *MIT App Inventor*, como foi confirmado na seção de resultados.

A primeira dinâmica serviu para explicar os conceitos de Pensamento Computacional na resolução de problemas e sua aplicação por meio de exemplos sem precisar do computador. O foco do projeto foi ressaltar aos alunos o investimento de tempo no processo de análise por meio do CT antes de desenvolverem algum aplicativo porque se eles conseguirem definir bem a lógica do problema, programar no *MIT App Inventor*, como em qualquer outra linguagem de programação, se torna uma tarefa mais simples.

No decorrer dos encontros, os alunos se familiarizaram com a interface gráfica do *MIT App Inventor* e o processo da criação da tela foi cada vez mais rápida, sem necessidade de muito acompanhamento da equipe e até alguns dos alunos personalizaram a aparência da tela de acordo a suas preferências. Na parte da programação em blocos, a maioria dos alunos, como foi apresentado na seção de resultados, conseguiu acompanhar, terminar e principalmente entender o que estava fazendo durante a montagem dos blocos.

Os aplicativos desenvolvidos na última dinâmica foram instalados e executados nos smartphones dos alunos, o que gerou uma maior motivação e satisfação dos alunos, haja vista que eles demonstravam para os outros alunos e para eles mesmos que tinham aprendido a criar aplicativos.

A metodologia do CT aplicada em cada um dos aplicativos desenvolvidos foi entendida pelos alunos e a maioria deles a consideraram uma boa estratégia, para resolver problemas não só no ambiente de computação, mas também que pode ser integrada à outras disciplinas. Os comentários recebidos pelos alunos da turma foram positivos: acharam o projeto produtivo para seu aprendizado de como resolver um problema e tinham muito interesse em continuar com mais aulas de *MIT App Inventor*.

A elevada adesão e motivação dos alunos aos princípios da CT e a facilidade no desenvolvimento de aplicativos, por meio do *MIT App Inventor*, confirmam que o uso do CT e desta ferramenta são positivamente significativas no aprendizado e são factíveis para inserção nas matrizes curriculares do ensino fundamental e médio do Brasil.

## 7. Referencias

- Rodriguez, C., Zem-Lopes, A.M., Marques, L., Isotani, S.: Pensamento Computacional: transformando ideias em jogos digitais usando o Scratch. An. do Work. Informática na Esc. 21, 62 (2015).
- Takahashi, T.: Sociedade da Informação no Brasil - Livro Verde. (2000).
- Prensky, M.: Digital Natives, Digital Immigrants. Horiz. 9, 1–6 (2001).
- Gomes, T.C.S., Melo, J.C.B. de: App Inventor for Android: Uma Nova Possibilidade para o Ensino de Lógica de Programação. An. dos Work. do Congr. Bras. Informática na Educ. 2, 620–629 (2013).
- Zaharija, G., Mladenović, S., Boljat, I.: Introducing basic Programming Concepts to Elementary School Children. Procedia - Soc. Behav. Sci. 106, 1576–1584 (2013).
- Grover, S., Pea, R.: Computational Thinking in K-12: A Review of the State of the Field. Educ. Res. 42, 38–43 (2013).
- Yadav, A., Zhou, N., Mayfield, C., Hambrusch, S., Korb, J.T.: Introducing computational thinking in education courses. Educ. Stud. 465–470 (2011).
- Almeida, M.E.B. de P., Brito, M.E.B.: Pedagogia de projetos e integração de mídias. Bol. 2003. (2003).
- Barr, V., Stephenson, C.: Bringing Computational Thinking to K-12: What is Involved and What is the Role of the Computer Science Education Community? ACM Inroads. 2, 48–54 (2011).
- Wing, J.M.: Computational Thinking. 49, 33–35 (2006).
- Costa, S.S., Autor2, 2, Autor3, 3, Autor4, 4: An exploratory study of games for introduction to computational thinking. 0–16 (2015).
- França, R.S. De, Tedesco, P.: Explorando o pensamento computacional no ensino médio : do design à avaliação de jogos digitais. 23o. Work. sobre Educ. em Comput. (2015).
- Gomes, T.C.S., Melo, J.C.B. de: O Pensamento Computacional no Ensino Médio : Uma Abordagem Blended Learning. XXI Work. sobre Educ. em Comput. - XXXIII Congr. da Soc. Bras. Comput. 651–660 (2013).
- Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J. a Y., Silverman, B., Kafai, Y.: Scratch: Programming for All. Commun. ACM. 52, 60–67 (2009).

---

# APÊNDICE D – ARTIGO: A SEVERITY BULLYING EPISODE IDENTIFIER FRAMEWORK BASED ON MACHINE LEARNING TECHNIQUES AND FUZZY LOGIC

---

O artigo “*A Severity Bullying Episode Identifier Framework Based on Machine Learning techniques and Fuzzy Logic*”, foi apresentado no ICAISC 2017<sup>23</sup> - “*16th International Conference on Artificial Intelligence and Soft Computing*” na cidade de Zakopane, Polônia o dia 12 de junho de 2017. Esse artigo forma parte do projeto de mestrado com a orientação do Prof. Dr. Edson Luiz Ursini e a coorientação do prof. Dr. Paulo Sérgio Martins.

## **A Severity Bullying Episode Identifier Framework Based on Machine Learning techniques and Fuzzy Logic**

Carmen Rosales Sedano<sup>1</sup>, Edson Luiz Ursini<sup>1</sup>, Paulo Sérgio Martins<sup>1</sup>

<sup>1</sup>University of Campinas, School of Technology, Limeira, Brazil  
c154133@g.unicamp.br, {ursini,paulo}@ft.unicamp.br

**Abstract.** *Bullying* at schools is a serious social phenomenon around the world that negatively affects the development of children. However, anti-bullying programs should not focus on labeling children as either bullies or victims since they could produce opposite effects. Thus, an approach to deal with bullying episodes, without labeling children, is to determine their severity, so that school staff may respond to them appropriately. Related work about computational techniques to fight against bullying showed promising results but they offer categorical information as a set of labels. This work proposes a framework to determine bullying severity in texts, composed by two parts: (1) evaluation of texts using Support Vector Machine (SVM) classifiers found in the literature, and (2) development of a Fuzzy Logic System that uses the outputs of SVM classifiers as its inputs to identify the bullying severity. Results show that it is necessary to improve the accuracy of SVM classifiers to determine the bullying severity through Fuzzy Logic.

**Keywords:** Machine Learning · Fuzzy Logic · Text Mining · Bullying

---

<sup>23</sup> ICAISC 2017 - <http://www.icaisc.eu/>

## 1. Introduction

Bullying is a serious social problem among school children with a high peak during early adolescence (11–14 years old) [1,2]. This problem concerns students, school staff and parents [3] because of its negative consequences for the mental, social and emotional student's health. The entire school environment becomes dangerous when there are not effective interventions against bullying situations and it involves all children, without exception [4]. For example, bullies are likely to suffer depression, to engage in fighting behavior as well as criminal and academic misconduct, to be physically and socially aggressive, to have low empathy and to have an exaggerated air of self-confidence [5]. On the other hand, their victims may suffer from depression, loneliness, anxiety, fear, sadness and low self-esteem [4]. Indeed, students who were either bullies or victims might have health problems and emotional adjustment in adulthood [6].

Therefore, it is important for schools to design and implement prevention and intervention programs to reduce bullying and victimization and the effectiveness of these programs promote positive youth [6]. Additionally, bullying should be viewed along a continuum rather than labeling children as bullies, victims or uninvolved, because labeling overemphasizes the role of individual children without accentuating their capacities and thus it contributes to a negative climate [7]. Therefore, instead of labeling children, an effective anti-bullying program has to be able to respond to signs of any of the categories of involvement [8]: bullying (helping bullies because their behavior is just a response to several causes such as familiar problems [1]) and victimization (helping victims, even by-standers, because typically they do not seek help or prefer to suffer in silence [7]). In [9], the authors propose a Bullying Assessment Matrix which allows school staff to determine the severity of a bullying episode. Indeed, that matrix would avoid labeling students and would help to measure the episode as a whole.

Typically, social scientists of bullying use surveys and questionnaires that may yield limited information due to their cost and potential participant fatigue [10]. In contrast, the computer science study of bullying is emerging with promising results [11]. Nowadays, digital technology forms part of the life of young people and the most prevalent online activity that they are engaged in is the use of social media such as Facebook, Twitter and YouTube. Digital technology may be exploited to determine bullying behavior [9] since the participants of a bullying episode (physical or cyber) often post social media text about their experience [11,12]. The work of Xu [12] shows that social media, with appropriate Machine Learning (ML) and Natural Language Processing (NLP) techniques, can be a valuable tool for the study of bullying.

Following that premise and the proposal of the Bullying Assessment Matrix of [9], this work proposes a bullying-severity identification framework composed by SVM classifiers developed in [12,13] and a Fuzzy Logic System (FLS) in order to determine the severity of bullying episodes occurring in texts.

The remainder of this paper is organized as follows: Sect. 2 presents literature review about bullying and current approaches to reduce bullying episodes at school. It also addresses work on computational techniques to fight against bullying. Section 3 describes our proposed approach used to develop the framework. Section 4 shows the experimental results and related discussion. Finally, Sect. 5 discusses the conclusions and future work.

## 2. Background and Literature Review

### 2.1. Bullying Definition

The most widely accepted definition of bullying was given by the psychologist Dan Olweus [2,14]. He defined that a student is being bullied or victimized when he or she is repeatedly exposed to negative actions on the part of one or more students over time. In fact, there are three key concepts [15] that differentiate bullying from other forms of school violence and conflict: (1) an intent to harm or upset another student, (2) the harmful behavior occurs repeatedly, and (3) the relationship between the bully (or bullies) and the victim(s) is characterized by an imbalance in power. In addition, each bullying episode has a severity level and an impact on the victim [11].

### 2.2. Types of Bullying

The forms of bullying behaviors are physical aggression (hitting, kicking, punching), verbal (name-calling, threatening), property damage (stealing or damaging the possessions of victims) [6], sexual (inappropriate touching, relational (spreading rumors or exclusionary behavior) and cyberbullying (bullying through electronic devices) [3,15,16]. Regarding the perception of bullying severity, physical bullying had higher values of severity than verbal and relational. However, verbal and relational bullying can be just as harmful as physical bullying [17].

### 2.3. Participants in a Bullying Episode

Typically, bullies and victims are the key participants in a bullying episode at school [14]. However, bullying is not an isolated event between two individuals [3]. In [3], we found the following roles: the bully, the victim and

three types of witnesses such as uninvolved students, bully supporters (students that incite the bully without personally taking action against the victim) and bully intervener (students that defend or console the victim). Additionally, the work of Xu et al. [13] augmented two new roles for social media: the reporter, who is someone that may not be present during the episode and the accuser, who accuses someone as being the bully.

#### 2.4. Bullying Assessment Matrix

In [9], the authors consider that all bullying episodes have to be responded appropriately with policies and processes according to their nature. Thus, they propose a Bullying Assessment Matrix to assess severity, impact and frequency of a bullying incident that supports schools in their decisions about how to treat the incident. The bullying incident is considered as moderate, major or severe if the total score varies between 3–5, 6–7 or 8–9 respectively.

#### 2.5. Prior Works about Bullying in Computer Science

Computational social science is an emerging field that has the capacity to collect and analyze large amounts of data, e.g. originating from the Internet [18] and especially from social media networks. Within this context, the computational study of bullying presented in [11–13] shows that social media can enrich it due to their nature (large-scale, near real-time and dynamic data) and their popularity among young people.

Indeed, the social network Twitter, that allows people to post 140-character messages called tweets, produces about 400 million tweets per day and has been used as a data source to answer social science questions [19,20]. Despite the fact that texts posted on Twitter may be unstructured and informal texts with noise as non-standard abbreviations, typographical errors, use of emoticons, irony and sarcasms [20,21], they offer the possibility to identify online (cyberbullying) and offline bullying trends (bullying episodes in the real world) [14].

In [22–25], we found Data Mining and ML techniques to detect cyberbullying using social networks as data sources, including Twitter. In contrast, the work of [10] propose the identification of participants of school bullying situations through their social interactions in a computer game instead of using real and social-interaction data from social networks - due to its evaluation complexity and its availability (e.g. Facebook). The thesis of [26] made a comparison between text classification methods and classification using sentiment analysis (emotional vectors) to determine bullying in Twitter conversations. The results showed that using the emotional vectors did not improve the accuracy of the classification (78% versus 75%).

#### 2.6. Support Vector Machine in Text Classification

In [12], there are five SVM classifiers publicly available, but we chose only four of them due to their importance according to the literature presented in Sect. 2. The work of [12] collected 32,477,558 tweets and each tweet was represented by the combination of both unigrams and bigrams (1g2g) which were part of a vocabulary file (4,524 entries). Each SVM classifier handles its own model file where we found that a weight was assigned to each feature of the vocabulary mentioned above. The description and accuracy of the chosen SVM classifiers are shown in Table 1. SVM classifiers in [12] return a real number. For SVM binary classifiers, results less than zero are labeled as No and greater than zero are Yes. For SVM multi-class classifiers, a real number is calculated for each possible class for a tweet, and the final result is the largest value among all of them.

#### 2.7. Fuzzy Logic System

A limitation that led to the creation of fuzzy logic was the poor precision offered by natural languages used to describe or share knowledge that is inherently vague for some contexts [27]. For example, concepts such as cheap, expensive, big and old do not have well-defined boundaries, and thus they may be called fuzzy concepts. The Fuzzy Logic (FL) approach allows us to better deal with the uncertainty present in problems with high complexity, where traditional system modeling techniques do not offer enough precision [27]. In Sect. 1, we have presented the complexity of the bullying problem and we have mentioned that labeling children would not be the correct path to solve the problem. Furthermore, Barton [3] shows that bullies and victims should not be completely (or precisely) regarded as such since it depends on the level of conflict in which they are involved in. Therefore, we propose to include and integrate the capabilities of Fuzzy Logic to determine the severity of bullying episodes according to a number of factors that would be inputs to our FL System.

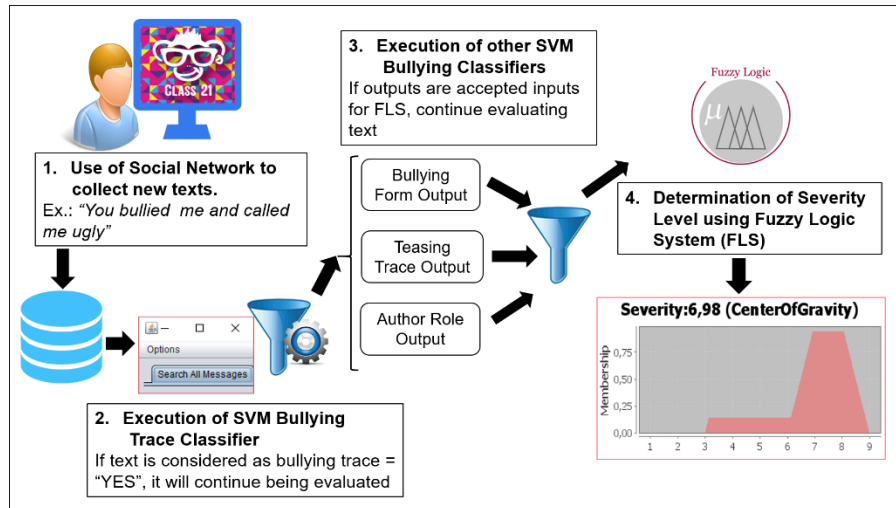
**Table 2.** SVM classifiers developed by [12]

SVM Classif.	Description	Classes	Accuracy (%)
Bullying Trace	It identifies if text belongs to bullying context	Yes No	86%
Teasing Trace	It identifies if exists lack of severity of a bullying episode.	Yes No	89%
Author Role	It identifies participants of bullying episode.	Victim, Defender, Reporter, Bully, Accuser, Other	61%
Bullying Form	It identifies bullying forms of bullying episode.	General, Cyberbullying, Physical, Verbal	70%

### 3. Proposed Approach

Our proposal consists in developing a framework that allows school staff to determine the severity level of a bullying episode in texts, in English written by students aged between 11 and 14 years old. The texts are collected through a social network developed for this project. They are evaluated using a Java Swing application that executes the SVM classifiers and the Fuzzy Logic System in order to obtain the severity level of the text. Figure 1 shows the functionality of the framework.

We divided the development of the framework in four main tasks: (1) Text pre-processing, (2) Text Classification using SVM classifiers of Xu [12], (3) Development of the Fuzzy Logic System, and (4) Development of a Social Network and Java Swing Application.

**Fig. 1.** Functionality of the Severity Bullying Episode Identifier Framework

#### 3.1. Task 1: Text Pre-processing

We collected 18,504 tweets using Twitter Streaming API and Tweepy from June to December 2016, and we also added 4,783 of the tweets used by Xu [12] in the training phase of the SVM classifier. These tweets included any of the keywords that belong to a bullying context according to the literature. Additionally, we collected texts from No-Place4hate.org and StampOutBullying.co.uk, which are stories written by children that were involved in bullying episodes. We analyzed tweets and stories to find high-frequency words, unigrams and bigrams, so that they could assist us in improving the vocabulary used by the SVM classifiers of [12]. We used only tweets (18,504 in total) to test the SVM classifiers in the next task. Since, tweets have an unstructured format, they passed through the Enrichment and Tokenization process defined by Xu [12]. We made some improvements in the Tokenization

process such as replacing URLs that started with https:// by HTTPLINK and correctly removing English contractions of words (e.g. he'll → he will).

### 3.2. Task 2: Text Classification using SVM classifiers available in the literature

To the SVM classifiers of [12], which were developed in Java, we added the following modifications: (1) created a method to read tweets collected in Task 1, (2) created an order of execution of SVM classifiers: It consisted on executing first the SVM Bullying Trace. Only tweets classified as YES can continue being classified by the other SVM classifiers. This scheme allows us to filter out those tweets that are not considered as bullying trace. For example, the tweet I was bullied in elementary school because of my height and then teased for crying about it I didn't have close friends until 3rd grade, that passed the filter, was classified with the following labels: general bullying form, No teasing and victim author role. We found that labels are associated with a real numeric value and that number varied in a range depending on the SVM classifier.

### 3.3. Task 3: Development of the Fuzzy Logic System

We followed an adaptation of the methodology proposed in [28] to develop the Fuzzy Logic System. Thus, we defined two parts: (1) Structure Identification, and (2) Parameter Identification.

#### Structure Identification

The structure identification is composed of (1) selection of the most relevant or possible input candidates and assignment of membership functions, (2) Specifications of the relationships between input and outputs through rules.

*Input selection and membership:* Once all tweets had their labels assigned due to the execution of the SVM classifiers, we analyzed them. Finally, we selected eight labels as input variables and we grouped them by three categories: author role, bullying form, and teasing. Only the Bully and Victim labels were considered out of the six categories defined in the literature: roles such as Reporter and Assistant were discarded due to their ambiguity; and Accuser and Defender labels were discarded due to the complexity involved in inferring their impact on a bullying episode. All of the labels of the bullying form (i.e. general, cyberbullying, verbal and physical) were considered due to their relevance according to the literature. The variable teasing was considered important because it allows determination of whether or not the text is a joke - and if so, the severity must decrease.

Next, we defined the Fuzzification process, i.e. for each input variable, we defined its membership functions. Each input variable related to a bullying form (general, physical, cyberbullying, verbal) and author role (bully, victim) had three membership functions: Low, Medium and High. Each input variable related to teasing had two membership functions: Yes and No. The intervals of each membership functions were determined based on the analysis of a set of 18,504 tweets labeled in the previous task. The output variable was called severity and had three membership functions: moderate, major and severe. We used the same names of the Bullying Matrix Assessment [9].

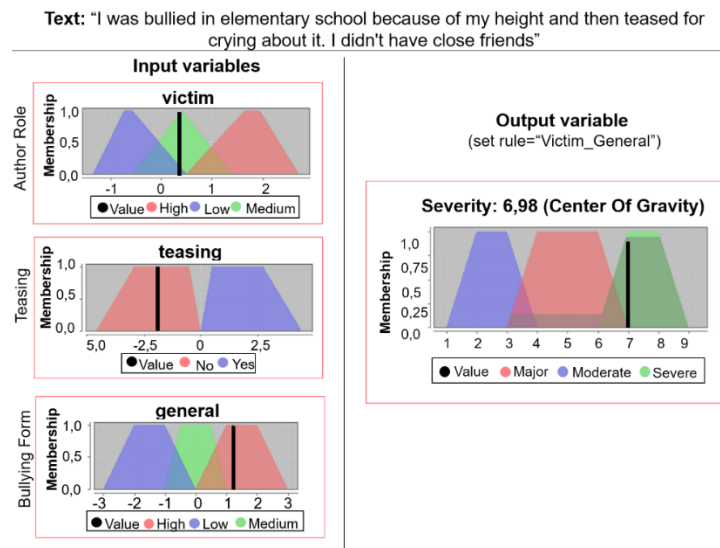


Fig. 2. Determination of the severity of bullying episode in a text using Fuzzy Logic



### *Rule generation*

Based on the literature review, we established some criteria for the creation of rules - for example - the importance of each input variable in determining the output variable severity. Thus, a higher importance was assigned to the input variable physical in comparison to the verbal, general and cyberbullying variables. Regarding roles, victim and bully were considered the most important ones to define the bullying severity. Finally, the “teasing” variable was considered most important when it belongs to the no category in its membership function. We generated eight sets of rules that were modeled in MATLAB software using the Mamdani fuzzy inference system: VictimGeneral, VictimPhysical, VictimVerbal, VictimCyberbullying, BullyGeneral, BullyPhysical, BullyVerbal, BullyCyberbullying. Each set has 18 rules, the set that will be used must be related to the labels of the tweet. For example, in Fig. 2, the tweet was evaluated with the set of rules VictimGeneral because of its labels (author = victim, bullying form = general and teasing=no) and the output was of severity severe.

### **Parameter Identification**

*Membership function tuning:* The analysis of labeled tweets allowed us to choose the shape of the membership function for each input and output variable. The range of each input variable was determined based on maximum and minimum values extracted from a set of 18,504 tweets.

*Parameter adjustment:* In [27], we found seven methods used for defuzzification: among them, the Centroid (or Center of Gravity) method is the most widely used. We have also decided to use it because of its accuracy.

### **3.4. Task 4: Development of a Social Network and Java Swing Application**

We developed a social network, in order to collect new texts, with the 140-character limitation - since there is no guarantee that the application of SVM classifiers works with larger texts [12]. This social network was developed in Wordpress (CMS), written in PHP and it uses MySQL as the database. To integrate the SVM classifiers with the Fuzzy Logic System, we used the jFuzzy-Logic.jar library in our Java Swing Application, which may be used by school staff to manage texts written by their students.

## **4. Results and Discussion**

A set of 18,504 tweets were collected in order to validate the modeled fuzzy logic system, and 11,004 were classified as bullying traces. Once the model processed the tweets, 10,717 were classified as general bullying, 28 as physical bullying, 86 as verbal bullying and 173 as cyberbullying. Finally, 5,949 tweets were classified with the variables of interest, i.e. the variables selected as inputs to the fuzzy logic system.

The resulting tweets were analyzed and the severity of the bullying episodes was determined by agreement of the authors, allowing the comparison to the output of the fuzzy system. The process revealed the difficulty in dealing with the identification of the severity of bullying episodes, even by common sense. The task of determining the severity on a bullying episode turned difficult as each author had a different perception of the same case, even when the same criteria used to create the fuzzy rules were applied.

## **5. Conclusions**

The accuracy of the SVM classifiers used in this work compromised the performance of the fuzzy logic system. However, it is important to emphasize the high complexity involved in addressing social issues from the field of computer science. The analysis performed in this work indicated that the accuracy of the outputs of the SVM classifiers directly affected the fuzzy outputs, meaning that, in order to obtain better results, the SVM classifiers need to be improved.

Currently, we continue working on improving the text classification in order to generate new models used by SVM classifiers. Other factors such as sex, race, social and economic status of the participants of bullying episodes should be considered as well as part of the evaluation of bullying episodes. In future work, other NLP techniques may be used in the preprocessing of texts. Furthermore, improvements in the vocabulary used by the classifiers are also recommended, for example, by assigning proper weights to the most relevant words pertaining to a given bullying context.

## **References**

1. Tusinski, K.: *The Causes and Consequences of Bullying*. (2008).
2. Bauer, N.S., Lozano, P., Rivara, F.P.: *The effectiveness of the Olweus Bullying Prevention Program in public middle schools: a controlled trial*. *J. Adolesc. Health*. 40, 266–274 (2007).

3. Barton, E.A.: The Bully, Victim, and Witness Relationship Defined. In: Bully prevention: Tips and strategies for school leaders and classroom teachers. pp. 1–18 (2006).
4. Lopes Neto, A.A.: Bullying: comportamento agressivo entre estudantes. *J. Pediatr. (Rio. J.)*. 81, 164–172 (2005).
5. Dake, Joseph A; Price, James H; Telljohann, S.K.: The nature and extent of bullying at school. *J. Sch. Health*. 73, 173–180 (2003).
6. Evans, C.B.R., Fraser, M.W., Cotter, K.L.: The effectiveness of school-based bullying prevention programs: A systematic review. *Aggress. Violent Behav.* 19, 532–544 (2014).
7. Ragozzino, K., Utne O'Brien, M.: Social and Emotional Learning and Bullying Prevention. (2009).
8. Rettew, D.C., Pawlowski, S.: Bullying. *Child Adolesc. Psychiatr. Clin. N. Am.* 25, 235–242 (2016).
9. Bullying Prevention Advisory Group: Bullying prevention and response : A guide for Schools. (2015).
10. Mancilla-Caceres, J.F., Pu, W., Amir, E., Espelage, D.: A Computer-in-the-Loop Approach for Detecting Bullies in the Classroom. In: Proceedings of the 5th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction. pp. 139–146 (2012).
11. Xu, J.-M., Zhu, X., Bellmore, A.: Fast learning for sentiment analysis on bullying. *Proc. First Int. Work. Issues Sentim. Discov. Opin. Min. - WISDOM '12*. 1–6 (2012).
12. Xu, J.: Understanding and Fighting Bullying with Machine Learning, (2015).
13. Xu, J.-M., Jun, K.-S., Zhu, X., Bellmore, A.: Learning from Bullying Traces in Social Media. 656–666 (2012).
14. Olweus, D.: A Profile of Bullying at School. *Educ. Leadersh.* 60, 12–17 (2003).
15. Dulmus, C.N., Sowers, K.M., Theriot, M.T.: Prevalence and Bullying Experiences of Victims and Victims Who Become Bullies (Bully-Victims) at Rural Schools. *Vict. Offender*. 1, 15–31 (2006).
16. California Department of Education: Bullying at School. (2003).
17. Pérez, V.: Percepción de Gravedad, Empatía y Disposición a Intervenir en Situaciones de Bullying Físico, Verbal y Relacional en Profesores de 5° a 8° Básico. In: *Psykhe (Santiago)*. pp. 25–37 (2011).
18. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Life in the network: the coming age of computational social science. *Science (80- )*. 323, 721–723 (2009).
19. Eisenstein, J., O'Connor, B., Smith, N. a., Xing, E.P.: A latent variable model for geographic lexical variation. *Proc. 2010 Conf. Empir. Methods Nat. Lang. Process.* 1277–1287 (2010).
20. Ajao, O., Hong, J., Liu, W.: A survey of location inference techniques on Twitter. *J. Inf. Sci.* 41, 855–864 (2015).
21. Ritter, A., Clark, S., Etzioni, O.: Named Entity Recognition in Tweets : An Experimental Study. *Conf. Empir. Methods Nat. Lang. Process.* 1524–1534 (2011).
22. Nalini, K., Sheela, L.J.: Classification of Tweets Using Text Classifier to Detect Cyber Bullying. In: *Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2*. pp. 637–645 (2015).
23. Kontostathis, A., Reynolds, K., Garron, A., Edwards, L.: Detecting Cyberbullying: Query Terms and Techniques. In: *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13*. pp. 195–204. ACM Press, New York, New York, USA (2013).
24. Dinakar, K., Reichart, R., Lieberman, H.: Modeling the Detection of Textual Cyberbullying. *Assoc. Adv. Artif. Intell.* 11–17 (2011).
25. Parime, S., Suri, V.: Cyberbullying detection and prevention: Data mining and psychological perspective. *2014 Int. Conf. Circuits, Power Comput. Technol. ICCPCT 2014*. 1541–1547 (2014).
26. Arthur Patch, J.: Detecting Bullying on Twitter using Emotion Lexicons, (2015).
27. Sivanandam, S.N., Sumathi, S., Deepa, S.N.: Introduction to Fuzzy Logic using MATLAB. Springer Berlin Heidelberg, Berlin, Heidelberg (2007).
28. Emami, M.R., Türksen, I.B., Goldenberg, A.A.: Development of a systematic methodology of fuzzy logic modeling. *IEEE Trans. Fuzzy Syst.* 6, 346–361 (1998).
29. Xu, J., Jun, K., Zhu, X., Bellmore, A.: Learning from Bullying Traces in Social Media. In: *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. pp. 656–666 (2012).
30. Afonso Mazon, J.: FIPE - Relatório Analítico Final. , São Paulo (2009).