UNIVERSIDADE FEDERAL DO CEARÁ

CAMPUS DE QUIXADÁ

CURSO DE ENGENHARIA DE SOFTWARE

MÁRIO SÉRGIO RODRIGUES FALCÃO JÚNIOR

ANÁLISE COGNITIVA PARA PROTEÇÃO DA CRIANÇA NAS REDES SOCIAIS

QUIXADÁ

JUNHO 2015

MÁRIO SÉRGIO RODRIGUES FALCÃO JÚNIOR

ANÁLISE COGNITIVA PARA PROTEÇÃO DA CRIANÇA NAS REDES SOCIAIS

Trabalho de Conclusão de Curso submetido à Coordenação do Curso Bacharelado em Engenharia de Software da Universidade Federal do Ceará como requisito parcial para obtenção do grau de Bacharel.

Área de concentração: computação

Orientador: Prof. Dr. Marcos Antônio de Oliveira

QUIXADÁ

2015

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca do Campus de Quixadá

F163a Falcão Júnior, Mário Sérgio Rodrigues
Análise cognitiva para proteção da criança nas redes sociais / Mário Sérgio Rodrigues Falcão Júnior. – 2015.
49 f. : il. color., enc. ; 30 cm.

Monografia (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Bacharelado em Engenharia de Software, Quixadá, 2015.
Orientação: Prof. Dr. Marcos Antônio de Oliveira
Área de concentração: Computação

1. Redes sociais 2. Internet e crianças 3. Mineração de dados (Computação) I. Título.

CDD 303.483

MÁRIO SÉRGIO RODRIGUES FALCÃO JÚNIOR

ANÁLISE COGNITIVA PARA PROTEÇÃO DA CRIANÇA NAS REDES SOCIAIS

Trabalho de Conclusão de Curso submetido à Coordenação do Curso Bacharelado em Engenharia de Software da Universidade Federal do Ceará como requisito parcial para obtenção do grau de Bacharel.

Área de concentração: computação

Aprovado em: ___ / junho / 2015

BANCA EXAMINADORA

_____

Prof. Dr. Marcos Antônio de Oliveira
Universidade Federal do Ceará-UFC

_____

Prof. MSc. Ticiana Linhares Coelho da Silva
Universidade Federal do Ceará-UFC

_____

Prof. MSc. Enyo José Tavares Gonçalves
Universidade Federal do Ceará-UF

AGRADECIMENTOS

A Deus por ter me dado saúde e força para superar os desafios.

Aos meus amados pais, Mário Sérgio Falcão e Sílvia Barros Falcão, que com muito estímulo se esforçaram para me dar a melhor educação, sempre me motivando a trilhar bons caminhos.

A minha irmã Natália Barros Falcão que de forma especial e carinhosa me deu força e coragem, me apoiando, preocupando-se até com os problemas pessoais pelos quais passei.

Aos amigos e colegas, pelo incentivo e apoio constante.

A minha querida namorada Amanda Cavalcante, por toda paciência, compreensão e carinho, e por me ajudar muitas vezes a achar soluções quando elas pareciam não aparecer.

Ao meu orientador prof. Dr. Marcos Antônio, que acreditou em mim, ouviu minhas considerações partilhando comigo as suas idéias, conhecimento e experiências que sempre me motivaram. Quero expressar o meu reconhecimento e admiração pela sua competência profissional.

RESUMO

A internet é um sistema global que conecta bilhões de pessoas em todo o mundo, possibilitando diversas maneiras de interação e organização social. Redes sociais, como Facebook, MySpace, Twitter e entre outras, têm criado uma nova forma de socialização que proporciona boas experiências aos usuários, porém acabam expondo um grupo específico, as crianças. Em virtude do risco oferecido em distintas camadas da internet o objetivo desse trabalho é desenvolver uma ferramenta inteligente que contribua na luta contra a ação de indivíduos que oferecem risco para as crianças utilizadoras da rede social Facebook, com a utilização de técnicas de mineração de dados e Sistemas Multiagentes.

Palavras chave: Aliciamento sexual infantil. Sistemas Multiagentes. Mineração de dados. Análise cognitiva.

ABSTRACT

The internet is a global system that connect billions of people worldwide, enabling diverse modes of interaction and social organization. Social networks like Facebook, MySpace, and Twitter among others, have created a new form of socialization that provides good experiences to users, however they expose a specific group of children. Due to the risk offered in distinct layers of the internet, the aim of this work is to develop a smart tool that helps in the fight against the action of individuals that are possibly a risky for children, users of the social network Facebook, using data mining techniques and multi-agent systems.

Keywords: Child sex grooming. Multi-agent Systems. Data Mining. Cognitive analysis.

LISTA DE ILUSTRAÇÕES

SUMÁRIO

1 INTRODUÇÃO

A internet é uma rede mundial que interliga milhões de computadores em todo o mundo,–servindo como um grande fator de comunicação e integração social (BELLONI, 2001). Um grupo de aplicações para internet são construídas com base nos fundamentos ideológicos e tecnológicos da Web 2.0, e que permitem a criação e troca de Conteúdo Gerado pelo Utilizador (UCG) (KIETZMANN, 2011), ou seja, blogs, páginas de relacionamentos, chats, redes profissionais (Linkedin, Rede Trabalhar), redes comunitárias (redes sociais em bairros ou cidades), redes políticas e principalmente redes sociais eletrônicas como Facebook, Twitter, Google +, MySpace e entre outros formam o grande conjunto das mídias sociais (LEMIEUX, 2008).

É pertinente que em nossa sociedade contemporânea as pessoas estejam mais próximas da tecnologia, principalmente as crianças, que possuem habilidades diferentes das de antigamente, ou seja, enquanto um criança da década de 80 possuía maior facilidade para construir ou modelar um brinquedo, as crianças da geração atual possuem habilidades para lidar com a informática, devido ao convívio rotineiro com a mesma (BOMBONATTO, 2007).

A evolução tecnológica aliada ao crescente acesso à internet fez com que crianças e adolescentes tivessem contato com a internet logo nos seus primeiros anos de vida. De acordo com a utilização das redes sociais novos conceitos são assimilados, exatamente por se encontrarem em uma fase de desenvolvimento e formação psicológica. Em razão disso e da fragilidade e vulnerabilidade da criança quanto à sua capacidade de discernimento e tomada de decisões conscientes, indivíduos mal intencionados, por conhecerem tais comportamentos, podem aliciar menores com o objetivo de praticarem algum tipo de aliciamento, como por exemplo exploração sexual.

As redes sociais são organizações virtuais compostas por pessoas, que dão a oportunidade das mesmas se relacionarem com diferentes tipos de indivíduos (RABESCO, 2013). A facilidade e a mobilidade dessas redes proporcionam todos os tipos de interesse por parte dos usuários, fundamentado na alta acessibilidade e utilizando diversos tipos de dispositivos e não apenas computadores. Dispositivos móveis como *tablets, smartphones, smartwatches (relógios inteligentes), smart tvs e óculos inteligentes*, são exemplos de mobilidade e acessibilidade destas ferramentas tão utilizadas atualmente. A acessibilidade

proveniente desses dispositivos ampliam o acesso à internet para um público variado e consequentemente permitem uma maior frequência nas interações por meio da rede.

O Brasil, de acordo com a revista EXAME, é o quarto país do mundo em número de smartphones (GUIMARÃES, 2013). O modo como a tecnologia vem evoluindo permite maior acesso à dispositivos que possam se conectar à internet, e aumentam a inclusão digital no Brasil. A analista de Wall Street, Mary Meeker, diz que o Brasil só fica atrás de China, Estados Unidos e Japão sobre o número de dispositivos móveis, e que 77% dos usuários de internet móvel têm no acesso a redes sociais a sua principal atividade online (MEEKER, 2011).

A popularidade de sites de relacionamentos, blogs, galerias de fotos na internet, sites de compartilhamento de vídeos e outros sites de compartilhamento de conteúdo tem explodido, resultando em mais informações pessoais e opiniões sendo disponibilizadas com menos controle de acesso (SANTOS, 2010). Em virtude de que redes sociais são redes de compartilhamento, com o intuito de difundir tudo o que nela se apresenta, compreende-se a falta de ênfase na preocupação com o controle de acesso.

Alguns resultados obtidos por Gutiérrez e Vega (2013) com 147 adolescentes de 13 anos de idade, usuários do Facebook Chat, asseguraram que 64.1% desses adolescentes já optaram em conversar com pessoas desconhecidas, 53.1% testemunharam possuir fotos sexualizadas de outros amigos e 77.1% chegaram a compartilhar tais fotos pela internet. Entende-se que a exposição de tais conhecimentos põe em risco as informações das crianças e adolescentes, levando em consideração que aliciadores sexuais usam redes sociais para buscar vítimas (POULSEN, 2006).

Há um déficit de inspeção na internet dos pais para com seus filhos, muitas crianças ficam expostas a um mundo até então desconhecido e formado por distintos tipos de personalidades (PEREIRA, 2009). Uma pesquisa realizada pela empresa Minor Monitor afirma que aproximadamente 38% das crianças na rede social Facebook não possuem a idade permitida para sua utilização e 30% dos pais permitem a utilização sem supervisionamento (SILVIO, 2012).

Conforme a empresa de segurança online Kaspersky Lab, o Brasil ocupa o quinto lugar no ranking mundial de CyberCrimes, dentre os delitos a pornografia infantil se destaca entre os primeiros desse ranking. O ponto de partida para isso começa em sites de relacionamentos e principalmente redes sociais, como por exemplo o Facebook que atualmente é a rede social mais utilizada no país (HITWISE, 2014).

Para reduzir o déficit dessa inspeção é interessante dispor de ferramentas automatizadas que identifiquem padrões irregulares ou suspeitos no uso das redes sociais. Dessa maneira o grupo infantil de usuários dessas redes estaria menos vulnerável aos possíveis ataques.

A essência da ferramenta a ser desenvolvida é buscar velar pela dignidade de crianças usuárias de redes sociais que, por conta de sua situação de pessoas em desenvolvimento, estão sujeitos a possíveis ataques na internet (BUCKINGHAM, 2000). Antecipar possíveis eventos indesejados pode oferecer maior segurança ao público alvo da aplicação deste estudo e tranquilidade aos respectivos responsáveis.

Existem trabalhos recentes com a mesma temática, não obstante tais trabalhos se utilizam de abordagens diferentes para identificar o acesso malicioso as crianças e adolescentes. O intuito desse projeto é propor uma abordagem inovadora que venha a somar com as ferramentas disponíveis atualmente.

Na seção a seguir alguns trabalhos relacionados com o projeto serão melhores descritos e comparados para que assim se possa ter um melhor embasamento sobre o que foi desenvolvido nesse trabalho.

2 TRABALHOS RELACIONADOS

A seguir alguns trabalhos relacionados ao tema serão apresentados de maneira breve e geral para que o leitor possa ter uma melhor compreensão das abordagens utilizadas.

2.1 Análise automática de textos em mensagens instantâneas para detecção de aliciamento sexual infantil

Em um estudo feito por Santin (2011), um serviço de software foi desenvolvido para classificar estágios de conversações em salas de bate papo, através de um conjunto de palavras pré-selecionadas e por um conjunto de regras.

O estudo de Santin (2011) utiliza o algoritmo SVM (Support Vector Machine) para classificar os estágios de interação entre as entidades, crianças e possíveis suspeitos, um dos quais também foi utilizado nesse trabalho. Outro método para gerar o modelo utilizado no estudo de Santin (2011), foi o uso da base de dados "www.perverted-justice.com" com conversações reais entre pedofílos e crianças. Essa mesma base de dados serviu de ajuda para coleta de dados e definição de heurísticas na construção do modelo responsável em classificar o risco da criança nas redes sociais neste trabalho. Entretanto a base de dados disponibilizada no site "www.perverted-justice.com" não foi utilizada diretamente, apenas serviu para o aprendizado de como os aliciadores sexuais interagem com suas vítimas e as persuadem.

Inconvenientemente, as palavras do bate papo deveriam ser exatamente iguais ao conjunto de palavras na base de dados, o que restringe bastante a precisão para identificar os estágios, já que as palavras podem variar de região para região ou de pessoa para pessoa. O tipo de comunicação em sites de relacionamentos chega a ser quase coloquial, composto por gírias, vícios da linguagem e expressões regionais que possivelmente podem não estar no conjunto pré-selecionado. Assim, o potencial para atingir uma grande massa de dados é muito restrito.

A principal diferença desse trabalho é a de analisar eventos que sobretudo não se alteram e não dependem de tanta acurácia quanto a de analisar uma determinada cadeia de caracteres, não apenas focando em uma análise literal como feito no trabalho de análise automática de textos feito por Santin(2011). Na abordagem deste trabalho busca-se utilizar também alguns comportamentos dos usuários na rede social para obtenção de conhecimento sobre suas ações e objetivos. Devido a quantidade de eventos do Facebook como curtidas, compartilhamentos, postagens, cutucadas, interesses, status de relacionamento, músicas, entre

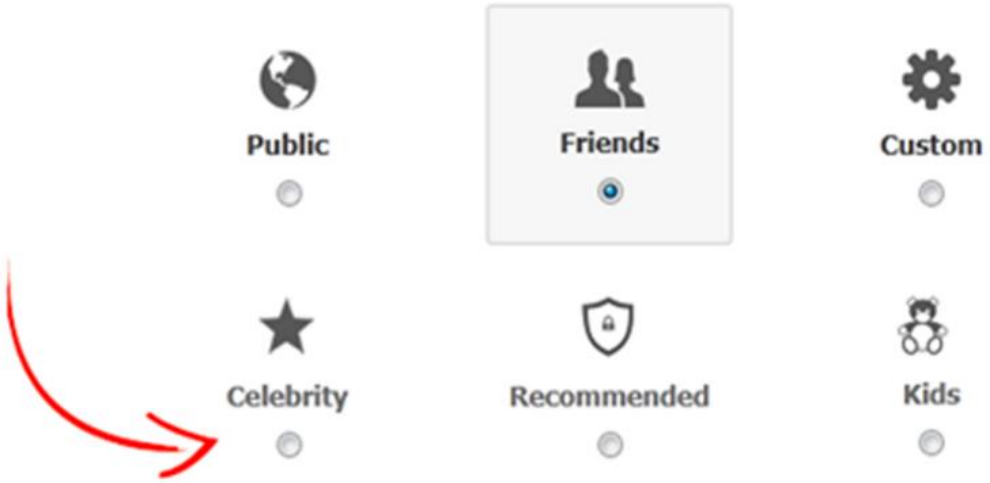
outros, o trabalho foi além apenas da análise de “texto”. Uma medida poderá ser retornada indicando algumas intenções dos usuário com quem a criança se relaciona na rede.

2.2 Identificação de perfis falsos nas redes sociais

A ferramenta Social Privacy Protector software for Facebook (SPP), tem por objetivo identificar perfis “falsos” na rede social Facebook, e melhorar as configurações de privacidade e segurança dos usuários (MICHAEL et al., 2012).

O SPP possui três camadas de proteção que aperfeiçoam a privacidade do usuário por meio da implementação de diferentes métodos. O sistema primeiro indica um possível perfil que pode representar uma ameaça, e logo em seguida fornece os meios para restringir suas informações pessoais para com o perfil suspeito. Em seguida a segunda camada permite ao dono do perfil na rede social ajustar suas configurações de privacidade de acordo com seu tipo de personalidade, por exemplo, na Figura 1 caso o indivíduo se considere uma “Celebridade” todas suas informações estarão dispostas ao público. Na situação “Recomendada” os dados estarão expostos apenas aos amigos. Já para o perfil “Crianças” apenas amigos terão contato e as requisições de amizades apenas estarão disponíveis para amigos de amigos.

Figura 1 - Configurações de privacidade



Fonte: Michael Fire (2013)

A terceira camada do sistema alerta ao usuário sobre a quantidade de aplicações instaladas nas contas de terceiros que possuem acesso aos seus dados privados. Tais dados,

podem possivelmente cair em mãos erradas e serem disponibilizados na internet sem o consentimento do titular (ELOVICI, 2012).

De forma análoga, a rede social utilizada na pesquisa do SPP é o Facebook, e o uso da Mineração de Dados para classificação das entidades também foi utilizado no desenvolvimento da ferramenta deste trabalho.

Entretanto, deve-se observar que o foco do SPP é identificar perfis "falsos", ou seja, pessoas que se passam por outras independente de seus objetivos finais. Analogamente, esse projeto busca identificar pessoas que ameacem um grupo infantil, de baixa faixa etária, sendo essas pessoas usuários "falsos" ou não. Esse grupo infantil faz referência as pessoas com até 12 anos de idade incompletos, de acordo com a lei 8.069, de 1990, do Estatuto da Criança e do Adolescente (ECA).

2.3 Proteção às crianças dos aliciadores sexuais onlines: Tecnologia, Psico-educacional, e considerações legais.

Um estudo feito por Dombrowski revela características gerais e estratégias utilizadas por aliciadores sexuais à medida em que estabelecem contato virtual com as crianças. Tais estratégias são utilizadas frequentemente por esses indivíduos e uma pesquisa nacional revela, que uma a cada cinco crianças, por meio da Internet, são solicitadas para estabelecerem relações sexuais anualmente (FINKELHOR at. el., 2001). Portanto, se faz necessária a constante vigilância por parte dos responsáveis pois o público alvo que mais sofre assédio são as crianças. A Internet mudou a maneira pela qual muitas pessoas interagem, ela agora é um tipo de serviço muito mais aceitável para a procura de amizades e relacionamentos românticos, especialmente entre gerações mais jovens (WOLAK, 2003).

As interações ocorrem de maneira gradual e podem variar de acordo com o momento, inicialmente podem começar com um contato offline, apenas mensagens como email ou chat, logo após conversações em tempo real, por meio de aplicativos ou até mesmo chamadas de celulares e por último face-a-face já introduzindo assuntos pornográficos e impróprios para o público infantil (YOUNG, 1997). Mais informações sobre as típicas atitudes dos aliciadores podem ser vistas na seção **4.4** deste trabalho.

O trabalho de Dombrowski propõe algumas sugestões de acordo com um ponto de vista tecnológico, psico-educacional e considerações legais. No domínio tecnológico alguns recursos podem ser abordados para proteger a criança, como por exemplo: a instalação de um *firewall*, prevendo acesso de terceiros e restringindo a visualização de informações privadas

instalação de um antivírus, instalação de um *key logger* para verificar todos os textos digitados na máquina local pela criança, monitoração do histórico do browser e um filtro de privacidade para restringir sites pornográficos e maléficos. Agregando maior valor na segurança contra os aliciadores, os recursos psico-educacionais irão somar de forma notória no combate aos aliciadores, dentre tais recursos podemos citar:

- Reconhecimento dos perigos, os pais caso encontrem devem dialogar com a criança e deixá-la ciente da possível ameaça em questão;
- Supervisão das amizades, muitas crianças estabelecem e mantêm amizades ao longo da Internet, algumas das quais podem ser prejudiciais e potencialmente nocivas (LONGO, 2000);
- Entender e aprovar o nome utilizado pelas crianças nas redes, predadores sexuais percorrem redes e podem atingir mais repetidamente aqueles jovens que utilizam nomes ou "nicks" com conotação sexual ou provocante (DEPARTAMENTO DE JUSTIÇA DOS EUA, 2001);
- Estabelecimento de um contrato entre o responsável e a criança, composto por uma baseline descrevendo regras e limites para se ter um controle maior, não total, sobre o uso da Internet feito pela criança;
- Alocar o computador em uma área pública da casa.

O terceiro recurso abordado no estudo são as considerações legais, ou seja, recorrer as leis disponíveis para apreender ou punir os aliciadores. Dessa forma a justiça será feita e a prevenção de futuros assédios sexuais poderá ser minimizada para quem usa a internet. A documentação pautando as mensagens, horários, recompensas oferecidas pelo aliciador, precisão sobre a sequência dos eventos, quantas pessoas tiveram conhecimento do processo poderá servir de estudo e investigação para que casos semelhantes possam ser prevenidos.

O trabalho de Dombrowski se relaciona com este devido as estratégias utilizadas pelos aliciadores sexuais para abordar uma criança. São padrões que foram implementados no modelo classificatório da ferramenta desenvolvida nesse trabalho, e além disso, a forma descrita no trabalho de como os responsáveis devem lidar com a criança e os conselhos psico-educacionais recomendáveis foram comentados com os responsáveis da criança no momento da emissão do relatório com a avaliação realizada pela ferramenta.

2.4 Intervenções para prevenir e reduzir abusos cibernéticos da juventude: Uma análise sistemática.

A Internet criou uma nova ferramenta de comunicação, em especial para os jovens, que utiliza-se de mensagens instantâneas, sites de redes sociais, YouTube, e-mails, salas e webCams, bate-papos e entre outros. Embora existam enormes benefícios que resultam nas comunicações eletrônicas, como base social e apoio acadêmico, a Internet é, no entanto, ao mesmo tempo um potencial local para abusos e vitimizações (WOLAK, 2003).

Com o intuito de evitar esses abusos cibernéticos algumas estratégias podem ser adotadas como tecnologias de software disponíveis para bloquear ou filtrar o acesso a esses conteúdos impróprios, intervenções dos responsáveis para lidar com as ameaças e terapias para as vítimas que já foram prejudicadas

O principal objetivo do trabalho de Mishna(2011) é fazer uma análise sistemática em estudos que combatem os abusos cibernéticos e medir a palpabilidade das técnicas para reduzir o risco nos comportamentos da criança.

Entre tais técnicas, a de se criar um grupo de controle para debater o risco que a internet pode oferecer e como se comportar para se distanciar dos aliciadores, desempenha papel importante para a segurança pessoal das crianças. Além disso a educação e conhecimento por parte dos responsáveis, pais e professores sobre os riscos que a internet pode oferecer pode proteger mais a criança contra os aliciamentos.

Essa pesquisa realizada por Mishna(2011) se fundamentou em mais de 3000 estudos, muitos dos quais foram irrelevantes, porém grande parte teve um significativo impacto nas conclusões.

As intervenções dos responsáveis no uso da rede social Facebook por parte da criança serviu de base nesse trabalho após o relatório obtido para lidar com a criança e ensinar-lhe seus respectivos limites. Os grupos de controle e orientação para os responsáveis da criança serviram de influências para aproximar ambos os lados, criança e responsável, a manterem conversações em busca de minimizar o risco oferecido pela internet. Isso pode se refletir na ferramenta a ser desenvolvida em relação ao relatório gerado que tem por objetivo ajudar a estabelecer as orientações entre os responsáveis e a criança.

Se nota que na pesquisa feita por Mishna(2011) os recursos tecnológicos que potencialmente podem ser utilizados para o combate dos aliciadores são definidos de maneira abstrata e pouco específica, por outro lado, nesse trabalho desenvolvido a descrição dos recursos tecnológicos é feita de forma detalhada e não exigirá por parte dos responsáveis um conhecimento tecnológico prévio de como utilizar a ferramenta ou usar os recursos. Dando para os responsáveis da criança resultados mais concretos e específicos.

2.5 Testando a eficácia de filtros da internet e decisões de bloqueio embutidas em quatro filtros web populares.

Desde que a Internet veio a atenção do público por volta de 1994, os americanos tornaram-se obcecados com o tão fácil acesso a pornografia, violência e discursos de ódio. Artigos de jornais e revistas têm alimentado esse medo com excitantes histórias sobre sites pornográficos, grupos de ódio, e predadores sexuais (TUROW, 1999).

Um dos recursos frequentemente utilizados são os filtros de internet que servem para bloquear determinados sites e conteúdos impróprios, infelizmente o estudo de Hunter afirma que alguns filtros conhecidos como CYBERsitter, Cyber Patrol, NetNanny e Surf Watch não demonstram tanta eficácia e deixam passar muitos acessos a conteúdos impróprios. O principal objetivo de seu estudo é utilizar um método científico e aleatório para identificar falhas na eficácia dos filtros de internet.

O método utilizado foi com a ferramenta Recreational Software Advisory Council's (RSAC) que categoriza um determinado contéudo como nocivo de acordo com os domínios: linguagem, nudez, sexo e violência.

Os filtros utilizados no estudo retornaram acesso negado apenas para as páginas que foram completamente categorizadas como nocivas pela ferramenta RSAC, porém os sites parcialmente classificados como perigosos, a grande maioria, não foram bloqueados pelos filtros. Os filtros classificaram os casos parciais apenas em 25%.

O trabalho desenvolvido buscou alcançar um percentual maior que o dos filtros em questão, utilizou de um modelo escalável e com boa acurácia para classificar os dados.

3 OBJETIVOS

Em razão de que a presença dos riscos para aliciamento sexual infantil é notória em diversas camadas da internet, principalmente nas redes sociais, o objetivo geral deste projeto é desenvolver uma ferramenta inteligente, composta por agentes autônomos, que analise os dados provenientes da rede social de uma criança e informe seu possível nível de vulnerabilidade à determinados riscos dentro da própria rede social.
Os objetivos específicos do trabalho envolvem:

1 - Recuperar dados específicos da criança sobre suas ações na Rede Social Facebook, em formato texto, para que assim possa ser realizado o processo de classificação das referências.

2 - Definir variáveis que influirão na classificação do grau de exposição e vulnerabilidade de crianças à ação de pedófilos e aliciadores, que usam o Facebook como forma de praticar crimes.
3 - Utilizar técnicas de Sistemas Multiagentes juntamente aos algoritmos de Mineração de dados.

4 FUNDAMENTAÇÃO TEÓRICA

4.1 Aliciamento Sexual Infantil implícito nas redes sociais

O aliciamento sexual infantil consiste em instigar crianças à prática do ato sexual com pessoas mais “velhas”**,** é uma forma de abuso infantil em que um adulto ou adolescente mais velho usa uma criança para estimulação sexual (HABIGZANG, 2005). Formas de abuso sexual infantil incluem pedir ou pressionar uma criança a se envolver em atividades sexuais independentemente do resultado, exposição indecente, órgãos genitais ou mamilos femininos, com a intenção de satisfazer os seus próprios desejos sexuais, intimidar ou aliciar a criança, ter contato físico sexual com ela, ou usá-la para produzir pornografia infantil.

Os efeitos do abuso sexual de crianças pode incluir depressão, transtorno de estresse pós-traumático, ansiedade, transtorno de estresse pós-traumático complexo, propensão a mais vitimização na idade adulta, lesão física, dentre outros problemas (AMAZARRAY,1998).

Segundo a lei, "abuso sexual infantil" é um termo guarda-chuva que descreve infrações penais e cíveis na qual um adulto se envolve em atividade sexual com um menor ou explora um menor para propósito de gratificação sexual. A Associação Psiquiátrica Americana afirma que "crianças não podem consentir em atividade sexual com adultos", e condena qualquer ação por um adulto: "Um adulto que se envolve em atividade sexual com uma criança está realizando um ato criminoso e imoral que nunca pode ser considerado como um comportamento normal ou aceitável socialmente ".

O aliciamento define a conduta social de um potencial agressor sexual infantil que procura ter alguma aceitação de suas investidas, por exemplo, em um chat. Todavia ultimamente se nota uma propagação de tais tentativas em redes sociais, já que os recursos oferecidos por esses serviços são superiores. No Facebook páginas de artistas infantis, Games, Grupos para crianças e entre outros são alvos acessíveis para esses indivíduos. Em razão disso não se pode deixar de afirmar que existem ameaças disfarçadas nesses grandes espaços de iteração virtual.

É relevante o uso das redes sociais como uma ferramenta aplicada para o aliciamento sexual infantil. Os pais ou responsáveis pelas crianças devem estar atentos e monitorando suas atividades na internet, e as ferramentas tecnológicas podem auxiliá-los a deduzir o que se passa no fluxo de dados entre a vítima e o potencial aliciador.

4.2 Sistemas Multiagentes

Sistemas Multiagentes (SMA) é uma sub-área da inteligência artificial distribuída e concentram-se no estudo de agentes autônomos, que tomam decisões próprias e se organizam dinamicamente, em um ambiente Multiagente. Um Sistema Multiagente (SMA) é um sistema composto por múltiplos agentes inteligentes que interagem entre si (MARIETTO, 2011). Podem ser utilizados para resolver problemas que são difíceis ou impossíveis de resolver para um agente individualmente. Sozinho o agente não coopera com outros agentes e sendo assim pode enfrentar maior dificuldade em resolver tarefas que exigem especialidades distintas e/ou maior processamento.

Alguns dos domínios em que o uso do SMA pode oferecer um enfoque adequado são: comércio eletrônico, resposta a desastres, modelo para estruturas sociais, games, transporte, logística, gráficos, sistemas de informações geográficas e entre outros.

O agente é uma entidade real ou virtual que está inserida em um ambiente podendo agir e se comunicar com outros agentes (OLIVEIRA, 2013). Pode possuir um comportamento autônomo e conhecimentos de domínio para aplicá-los a outros agentes. Funciona como uma abstração de um sujeito do mundo real, por exemplo, um agente de trânsito deve possuir características como:

1. Conhecimento: Leis e comportamentos dos condutores de veículos;
2. Objetivo: Fazer com que as leis sejam respeitadas;
3. Ações: Multar, apitar, parar e etc.

O ambiente em que esse agente de trânsito deverá trabalhar poderá ser uma cidade repleta de outros agentes condutores de veículos.

Entre algumas abordagens e utilizações para agentes podemos citar por exemplo: “Um agente é qualquer coisa que pode ser vista percebendo em um ambiente por meio de sensores e atuando no mesmo por meio de atuadores” (NORVIG, 2012), “O termo agente é utilizado para representar dois conceitos ortogonais. O primeiro é a habilidade de execução autônoma e o segundo é a habilidade em domínios específicos” (SANKAR, 2013).

Também chamado de sistema inteligente pois não deixa de ser um sistema distribuído em que todos os nodos são sistemas de inteligência artificial capazes de agir de maneira inteligente (DAMASCENO, 2012).

Alguns frameworks que implementam plataformas e protocolos de comunicação entre agentes surgiram para estabelecer padrões de desenvolvimento para SMA. Esses modelos

economizam tempo dos desenvolvedores e promovem a padronização nos SMAs desenvolvidos.

Os protocolos podem ser definidos como um conjunto de padrões de troca de mensagens e práticas especificadas para facilitar a interação entre os agentes. O uso pragmático desse recurso não é obrigatório, no entanto, caso seja aplicado, deverá seguir à risca as respectivas regras.

Os protocolos de comunicação são usualmente definidos em vários níveis (STEPHENS et al., 1999). Os níveis inferiores definem o método de interação dos agentes. Os níveis intermediários definem o formato (sintaxe) da informação transmitida. Nos níveis superiores encontram-se as especificações do sentido (semântica) da informação.

No que diz respeito à aridade, os protocolos de comunicação subdividem-se em protocolos de aridade binária e aridade n. Enquanto um protocolo binário, envolve apenas um emissor e um receptor, um protocolo de aridade n implica a existência de um emissor e múltiplos receptores. Genericamente, podemos definir que um protocolo contém a seguinte estrutura de dados (STEPHENS et al., 1999):

- Emissor;
- Receptor(es);
- Linguagem utilizada;
- Funções de codificação e descodificação da linguagem;
- Ações que o receptor deve executar.

Levando em consideração que essa estrutura possui relevância para padronização em qualquer tipo de interação definida por um protocolo de comunicação, os demais protocolos utilizados no mercado seguem esse padrão, inclusive o qual será utilizado nesse trabalho que procede da família dos protocolos FIPA (www.fipa.org), cujo principal objetivo é a definição de normas para tecnologias conceituais e computacionais usadas para a criação e operação de agentes e de SMAs (BELLIFEMINE, 2007).

4.3 Análise Cognitiva por meio da Mineração de Dados

O processo de análise cognitiva serve para obter conhecimento sobre algum domínio por meio de percepção, memória, raciocínio, juízo, imaginação, pensamento ou linguagem. De forma clara se pode dizer que a cognição é o arranjo com que o cérebro aprende e recorda por meio dos cinco sentidos (ROSE, 2012).

No âmbito desse trabalho qualquer perfil da rede social que represente uma ameaça foi classificado como inclinado a desenvolver algum tipo de relacionamento com o intuito de aliciar a vítima. Tal classificação foi verificada fazendo-se uso das técnicas de mineração de dados.

A mineração de dados é o processo da descoberta de informações que não estão explicitas em um grande conjunto de dados. Técnicas, padrões e algoritmos de aprendizagem são utilizados, empenhando-se em identificar tendências. Não se pode simplesmente notar tais valores apenas com a exploração tradicional, estratégias matemáticas se fazem necessárias para destacar os comportamentos (REZENDE, 2003).

Os relatórios obtidos com a mineração podem ser apresentados de diversas formas: agrupamentos, hipóteses, regras, árvores de decisão, grafos, ou dendrogramas (MTA, 2006).

A mineração de dados faz parte da rotina diária de diversas empresas que possuem recursos e se importam com o "Business Value" que as informações de seus clientes possuem e que os mesmos podem ter seus lucros aumentados de acordo com as tendências identificadas (AULAR, 2007). Devido tais negócios acumularem grande volume de dados em seus aplicativos operacionais, estatísticas podem ser utilizadas de maneira refinada buscando especificar padrões de consumo e motivações dos que utilizam os serviços. Dessa forma, possibilidades irão surgir para que a companhia possa fazer, priorizar ou mudar sua organização para que atenda de maneira mais rentável as solicitações do público.

Dentre as técnicas de mineração de dados existentes, há uma que melhor se encaixa nesse trabalho: a Classificação, que é responsável por reconhecer modelos que descrevem o grupo ao qual o item pertence por meio do exame dos itens já classificados e pela inferência de um conjunto de regras (MALUCELLI, 2010). Por exemplo, empresas de operadoras de cartões de crédito e companhias telefônicas preocupam-se com a perda de clientes regulares, a classificação pode ajudar a descobrir as características de clientes que provavelmente irão abandoná-las, e oferecer um modelo para ajudar os gerentes a prever essa situação, de modo que se elabore antecipadamente campanhas especiais para reter esses clientes.

De maneira análoga, os perfis dos usuários ameaçadores serão classificados e fundamentados em um modelo prévio definido que contenha características, aspectos e atributos do possível aliciador, como está definido a seguir na seção 4.4, bem como do nível de vulnerabilidade ao qual as crianças estão expostas ao fazerem uso do Facebook.

4.4 Modelo classificatório de um aliciador sexual

Alguns atributos devem ser definidos contendo os campos escolhidos na biblioteca Graph API do Facebook, melhor definida ao final da seção 6.1, e quais possíveis valores podem assumir para classificar o nível de exposição da criança a algum usuário aliciador.

Padrões, recursos e palavras chaves são frequentemente utilizados por aliciadores para conseguirem seus objetivos, como o envio de e-mails, recados em blogs, convites para sites de encontro, envio de imagens impróprias para crianças, utilização de aplicativos mobiles para paqueras, como por exemplo, Tinder, Flert, Cuddlr, Zoosk, OkCupid, Pof e dentre outros que possuem temáticas semelhantes (ALGERIS, 2006).

Determinados tipos de comportamentos dentro desses sistemas de relacionamentos virtuais são relevantes para induzir ao conhecimento das intenções por trás das ações feitas, em algum momento anterior, pelos usuários.

Compartilhamentos de segredos, participações em temas polêmicos, tipos de fotos adicionadas, associações em determinados grupos, games, tipos musicais e exposição ao extremo da vida privada, familiar e financeira podem ser temas significativos para avaliação e classificação do nível de suspeita que o perfil possui dentro da rede social (MIRANDA, 2000).

Há casos em que certas contas vão possuir informações falsas, por isso, a ausência de determinados dados também podem influenciar no nível de desconfiança do perfil. Caso o usuário não possua fotos pessoais ou de pessoas físicas reais, parentesco registrado, histórico de estudos ou trabalho, locais visitados em seus registros e dentre outros indícios de natureza semelhante, não se pode afirmar com toda propriedade que esse perfil corresponde a um usuário honesto sobre suas informações, todavia também não se pode afirmar que seja o perfil de um aliciador sexual infantil. No entanto são indícios que podem contribuir na classificação do usuário como um suspeito ou não.

Grande parte dos aliciadores sexuais infantis costumam seduzir gradualmente as crianças dando atenção, atuando de forma gentil e dando presentes (MIRANDA, 2000). Quando conquistam a confiança tentam distanciar a criança do contato familiar. Nas conversas, esses aliciadores potencializam problemas familiares como forma de falso apoio, aproximando-se ainda mais das vítimas, dedicam uma considerável parte do tempo à aproximação. O momento mais crítico é durante a noite. Conhecem as músicas da moda, hobbies e interesses da criança (MIRANDA, 2000). A partir do momento em que conquistam a confiança, começam a compartilhar material pornográfico de forma gradual nas conversas como forma de iniciar o contato sexual. A pornografia adulta é utilizada por essas pessoas

para criar a ilusão de que relacionamentos entre crianças e adultos é algo comum. O objetivo principal é o contato por vídeo para depois realizarem encontros pessoais (WORTLEY, 2012).

Predadores sexuais são um grupo heterogêneo e, como resultado, é difícil de definir uma tipologia do predador sexual. Historicamente, predadores sexuais têm sido retratados como homens mais velhos, da classe média (GUDJONSSON, 2000). Pesquisas recentes sugerem que os predadores sexuais são adultos na faixa etária de 18 à 72 anos, com predomínio entre 30 a 42 anos de idade (ELLIOTT, 1995). No entanto esse crime sexual não se limita a idade adulta. Muitos jovens também cometem atos de violência sexual (MIRANDA, 2000), no máximo 30% a 60% dos casos de abuso sexual infantil nos Estados Unidos são cometidos por crianças de idade inferior a 18 (DAVIS, 2002).

As estatísticas indicam que os aliciadores sexuais de todas as idades são predominantemente do sexo masculino, sendo responsável por 85% a 90% dos casos de abuso sexual (DAVIS, 2002). Quando estamos falando de um abuso sexual físico, o típico agressor sexual é muitas vezes alguém bem próximo da criança, ou seja, tio, amigo relativo, família, vizinho (CRESPI, 2002), por outro lado quando falamos de aliciamento sexual na internet os principais suspeitos não são os mais próximos da criança, e sim os distantes. Mitchell realizou um estudo e descobriu que quase 48% dos aliciadores sexuais da internet foram com idade inferior a 25 anos, e quase um quarto eram do sexo feminino. Além disso, 97% das solicitações onlines foram de estranhos e não de pessoas próximas à criança ou sua família (MITCHELL et al., 2001).

As crianças gostam de atenção e aceitação por parte das pessoas, e muitas vezes quando não encontram isso dentro de casa procuram na internet, e este é um dos pontos fracos que os aliciadores podem se aproveitar e agir, transparecendo confiança e devotando atenção às crianças acabam recebendo admiração pelas vítimas (SMITH, 1989). E a seguir poderá introduzir conversas com conotação sexual e realizar troca de fotos.

Frequentemente o artifício utilizado para as conversações e encontros virtuais são os chats, no entanto recentemente o uso das redes sociais e jogos online que possuem bate papo vem ganhando espaço como um recurso para o estabelecimento do contato inicial (FAVERO, 2014). Por esse fator é relevante que as crianças não compartilhem fotos ou informações pessoais, como nome, escola onde estuda e número de telefone.

O monitoramento por parte dos responsáveis da criança deve ser presente para evitar potenciais cenários desagradáveis.

4.5.Modelo classificatório da criança alvo de aliciadores sexuais na internet

Há certas características que podem predispor uma juventude para vitimização sexual. A literatura indica que as crianças com baixa auto-estima são mais visadas para o abuso sexual (ELLIOT et al., 1995). Mais tipicamente, os jovens que vêm de famílias desestruturadas e pobres são mais vítimas de violência sexual (KENNY, 2000). Isso ocorre porque famílias empobrecidas estão mais suscetíveis a terem problemas sociais de todos os tipos, incluindo abusos sexuais (SNOWDEN, 1999). Distúrbios emocionais, dificuldades escolares, síndrome de rejeição, desamparo e abertura emocional acabam tendo relevância para deixarem as crianças como alvos mais acessíveis aos aliciadores.

Aqueles que são mais vulneráveis à vitimização são menos propensos a ter pessoas responsáveis envolvidos ativamente em suas vidas (GLASSER et. al., 2001).

Desta forma se faz necessária a contínua presença dos responsáveis na vida da criança, seja dentro de casa, na escola, na roda de amigos e demais ambientes frequentados por ambos.

4.6 Algoritmo SVM e SMO

O SVM (Support Vector Machine) é um conceito na computação para um conjunto de métodos do aprendizado supervisionado que analisam os dados e reconhecem tendências, por isso pode ser usado para classificação e análise de regressão (CORTES, 1995).

A técnica SVM pertence a uma categoria de classificadores lineares e foi desenvolvida por Vapnik, com o objetivo de auxiliar na solução de problemas, não só de classificação, mas também no reconhecimento de padrões. O seu conceito é baseado na idéia de minimizar riscos estruturais, ou seja, minimizar erros da classificação empírica e maximizar a margem geométrica entre os resultados (WITTEN, 2005).

O algoritmo do WEKA responsável por implementar essa técnica SVM se chama SMO (Sequential Minimal Optimization), implementado por John Platt. Com a utilização desse algoritmo, a utilização de memória é linear para realizar os treinamentos. Com isso, o SMO permite lidar com grande quantidade de arquivos para treinamento (WITTEN, 2005).

4.7 Algoritmo J48

Esse algoritmo permite a criação de modelos de decisão em árvore. Utiliza uma metodologia gulosa para induzir árvores de decisão para posterior classificação (WITTEN, 2005). O modelo de árvore de decisão é construído pela análise dos dados de treino e o modelo utilizado para classificar dados ainda não classificados. O J48 gera árvores de decisão, em que cada nó da árvore avalia a existência ou significância de cada atributo individual. As árvores de decisão são construídas do topo para a base, através da escolha do atributo mais apropriado para cada situação. Uma vez escolhido o atributo, os dados de treino são divididos em sub-grupos, correspondendo aos diferentes valores dos atributos e o processo é repetido para cada sub-grupo até que uma grande parte dos atributos em cada sub-grupo pertençam a uma única classe. A indução por árvore de decisão é um algoritmo que habitualmente aprende um conjunto de regras com elevada acuidade (WITTEN, 2005).

4.8 Algoritmo IBK

O algoritmo IBK é uma versão do algoritmo de clusterização k-NN (k-nearest neighbor) utilizado em tarefas de clusterização. Esse método representa cada instância como um ponto de dado em um espaço d-dimensional, onde d é o número de atributos. Dada uma nova instância, calcula-se a sua proximidade com o resto dos pontos de dados no conjunto de treinamento, usando medidas de proximidade, tais como a distância euclidiana. Os k vizinhoos mais próximos de uma instância z são classificados como sendo da mesma classe de z (PINHEIRO, 2008).

4.9 Árvores de decisão

Árvores de decisão são similares a regras “Se-então”. É uma estrutura muito usada na implementação de sistemas especialistas e em problemas de classificação. As árvores de decisão tomam como entrada uma situação descrita por um conjunto de atributos e retorna uma decisão, que é o valor predizido para o valor de entrada (RYAN, 2004).

O modelo classificatório a ser gerado é estilo Bottom-up, ou seja, obtenção do modelo de classificação pela identificação de relacionamentos entre variáveis dependentes e independentes em bases de dados rotuladas (CHAPELLE, 2006). Isso deve ao fato de não haver uma base de dados com informações reais. É por esse motivo que as árvores de decisão se encaixarão adequadamente, porque elas implementam esse tipo de abordagem.

A classe definida se há risco para a criança ou não é uma variável dependente cujo valor é definido a partir das variáveis independentes, ou seja, a partir dos atributos.

Árvores de decisão são geralmente aplicadas junto a grandes bases de dados. Para tanto regularidades implícitas presentes na base de dados devem ser descobertas automaticamente e expressas, predominantemente, na forma de regras (CHAPELLE, 2006).

5 RELEVÂNCIA DO TRABALHO

É expressível o impacto das redes sociais sobre a sociedade e o quanto ela está presente em todas as classes sociais e faixas etárias. Segundo o provedor de internet UOL, o Brasil detêm mais de 76 milhões de usuários ativos no Facebook, o que significa aproximadamente um terço dos habitantes do país (REUTERS, 2014).

De acordo com um estudo feito pela VelocityDigital, 25% dos utilizadores do Facebook não tratam dos seus parâmetros de confidencialidade, levando isso em consideração, certos tipos de comportamentos e classificações de perfis podem ser feitos procurando detectar atos que possam ferir os direitos humanos (COOPER, 2013).

Na Figura 2 há um estudo feito pela Tic Kids Brasil sobre alguns percentuais do envolvimento das crianças e adolescentes nas redes sociais.

Figura 2 - Como os adolescentes usam internet no Brasil

Fonte: Edgard Matsuki

Como principal medida de segurança, os pais afirmam olhar o histórico da internet dos filhos. Aproximadamente 50% dos usuários de 9 a 16 anos afirmam saber apagar os sites visitados (MATSUKI, 2012). Além disso o aliciamento de menores não ocorre em sites pornográficos ou outros que possuam por domínio nomes como “www.aliciamosmenoresdeidade.com.br”. Em redes sociais e chats utilizados com frequência, o envolvimento com diversas pessoas por meio de conversações podem induzir a ocorrências de sedução e suborno.

Em razão disso, soluções computacionais focadas em detectar tais situações podem ajudar na prevenção da pedofilia e na proteção e defesa das vítimas. Consequentemente, a maneira como a criança interage nas redes sociais e na web de forma geral poderá ser monitorada e regulada por seus responsáveis, gerando maior confiança aos pais e segurança às crianças.

A aplicação da mineração de dados nesse cenário pode ter um impacto relevante em relação à classificação de perfis, entretanto é imprescindível o acesso aos dados dos usuários, nos quais em alguns casos em particular, não são muitos e podem dificultar o processo de classificação da criança para os algoritmos de mineração de dados devido a escassez de informações (WITTEN, 2005).

6 PROCEDIMENTOS METODOLÓGICOS

6.1 Coleta dos dados da rede social Facebook

Para que haja um público a ser investigado se faz necessária a existência de conteúdo. Em virtude de que a rede social Facebook possui um grande volume de dados e distintos tipos de perfis de usuários alguns atributos desempenham papel importante na interpretação de sentimentos, captura de costumes, gostos e entre outros. Em virtude disso algumas características serão relevantes para análise e decisão dos possíveis riscos oferecidos à conta da criança, entre esses atributos podemos citar: feed de notícias, interesses, músicas, vídeos, páginas, enlaces curtidos ou compartilhados, biotipo, data de nascimento, inspirações, tipo de educação, religião, conversações, postagens, datas e horários impróprios de atividades registradas na rede social, jogos, eventos, canais de televisão, família, álbuns compartilhados e com excesso de fotos, comentários maldosos em fotos e postagens, mensagens privadas de amigos ou terceiros que contenham conotação sexual, nota de privacidade que informa o quão exposto as informações pessoais estão e amigos que possam representar uma ameaça.

O perfil em questão a ser investigado será somente o da criança, até 12 anos de idade de acordo com o estatuto da criança e do adolescente no Artigo 2º. Obrigatoriamente os pais devem conceder permissão a aplicação para ler suas informações públicas e privadas.

Os dados do Facebook são disponibilizados aos desenvolvedores registrados, assim, se torna factível utilizá-los para o sistema a ser desenvolvido. A interface de programação da aplicação (API) para estabelecer comunicação com o servidor do Facebook será a Graph API, uma biblioteca que concede permissões para ler e escrever na base de dados da rede social.

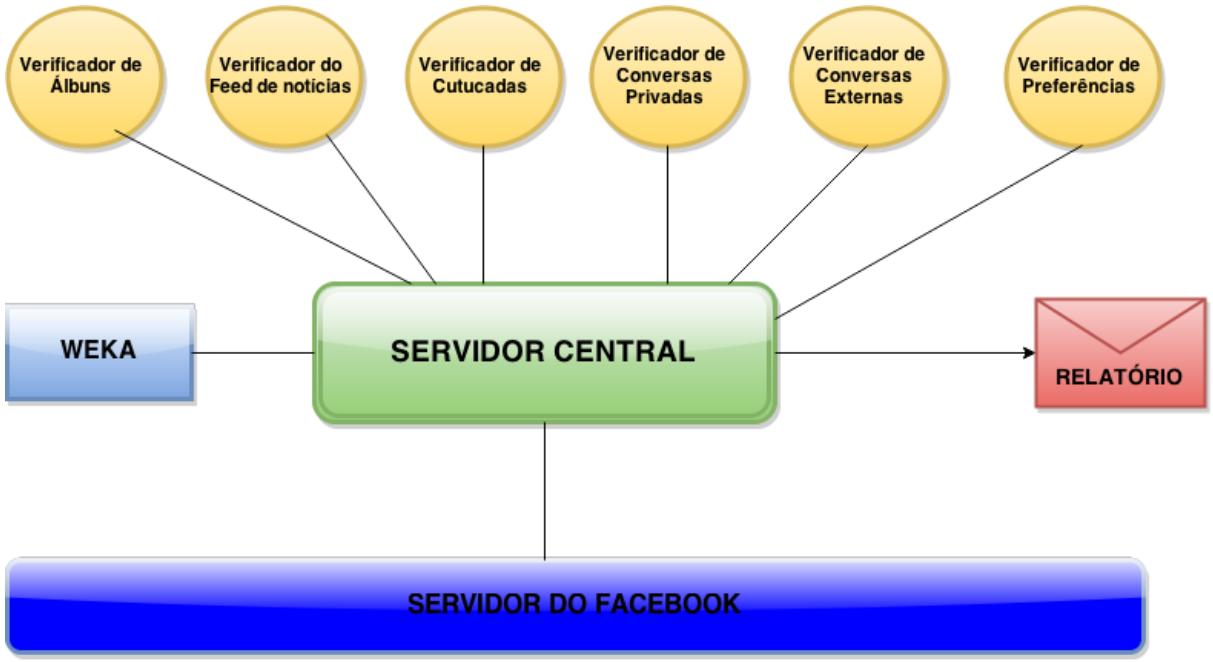
Serão criados módulos iniciais do sistema que sejam responsáveis por requisitar os dados no servidor do Facebook através de um arquivo JSON e após isso enviá-los para processamento aos módulos de mineração de dados e sistemas Multiagentes.

6.2 Desenvolvimento do sistema

O sistema desenvolvido é constituído por Agentes Autônomos desenvolvidos no framework Jade (Java Agent Development Framework), e utiliza dos protocolos padrão FIPA (Foundation for Intelligent Physical Agents), como descrito na seção 4.2, para comunicação entre os agentes.

A arquitetura utilizada para comunicação entre os agentes, servidor fornecedor de dados do Facebook e o módulo de mineração de dados do Weka pode ser melhor visualizada na Figura 3, onde se pode notar a interação entres os módulos. Após toda comunicação entre os agentes autônomos (marcados de amarelo na Figura 3), tomadas de decisões e processamento o servidor central retornará para o usuário o nível de vulnerabilidade e exposição ao qual a criança está sujeita.

Figura 3 - Arquitetura do sistema



Fonte: Arquitetura feita na ferramenta UML online

Os agentes autônomos interagem entre si recolhendo dados, definidos na seção 6.1, do servidor do Facebook e filtrando essas informações, trabalham em conjunto até gerar um objeto “Criança”, que possui referências a informações consideradas relevantes para o modulo de mineração de dados, a seguir este último tem por responsabilidade receber esse objeto “Criança” por meio de uma chamada de método e utilizar dos algoritmos de classificação para categorizar essa instância através do modelo de classificação do próprio sistema retornando ao servidor central o nível de vulnerabilidade ao qual a criança está suscetível.

O objeto criança é formado por campos que foram preenchidos com os resultados das análises feitas pelos agentes autônomos que servirão para serem enviados ao módulo de mineração de dados, são estes os campos:

- Postagens: Um campo numérico que informa a segurança das postagens feitas pela criança e as quais ela foi marcada. Qualquer valor acima de 3.0 já é considerado como não seguro. Esse campo numérico é calculado a partir da média ponderada:

Postagens *= ((quantidade de datas perigosas) * 3 + (nota de privacidade * 1) + (quantidade de mensagens perigosas * 2) + (quantidade de amigos perigosos * 2) + (quantidade de descrições e comentários perigosos das postagens *2)) / 10*

- Álbuns: Um campo booleano que recebe "Verdadeiro" ou "Falso", caso seja verdadeiro a criança não corre risco nas descrições, quantidade de fotos de seus álbuns e alguns compartilhados com outros amigos, do contrário ele está sujeita a alguns riscos.
- Preferências: O campo preferências é constituído por um conjunto de valores, notas nos itens: livros, músicas, filmes, eventos (exemplo: festa, show, aniversário e etc), canais de televisão e grupos vinculados a sua conta, após uma média ponderada nesses itens o valor de preferências é apontado como "Verde", "Amarelo" ou "Vermelho" em ordem crescente de perigo. Esse campo numérico é calculado da seguinte forma:

Preferências *= ((quantidade de eventos perigosos * 1) + (quantidade de grupos perigosos * 3) + (quantidade de livros perigosos * 2) + (quantidade de filmes perigosos * 3) + (quantidade de músicas perigosas * 3) + (quantidade de canais de televisão perigosos * 2)) / 14.*

- Feed de notícias: Um campo booleano que recebe "Verdadeiro" ou "Falso", caso palavras de teor pejorativo e sexual ou conteúdo não indicados para o menor sejam observados, por exemplo: violência, abuso sexual, forró, safada, xvideo, redtube, pelada, mande uma foto, transa, masturbação, erotismo, segredo, aviões, vou contar pra sua mãe, pornô, carnaval, folia, dentre outras mais explícitas.
- Cutucadas: Um campo numérico que informa a quantidade de cutucadas recebidas pela criança durante a madrugada (00:00 até 06:00).

- Família: Um campo booleano que recebe “Verdadeiro” ou “Falso” informando se a família da criança está cadastrada na sua conta ou não, se estiver, o campo receberá “Verdadeiro”, do contrário receberá “Falso”.
- Jogos: Um campo numérico que informa a quantidade de jogos não aconselháveis utilizados frequentemente pela criança, por exemplo, tinder, cupid, interesting, skout, let’s date, meetmoi e catra. Qualquer valor acima de 5 já é considerado como não seguro.
- Vídeos: Um campo numérico que informa a quantidade de vídeos não aconselháveis vistos pela criança. Os vídeos considerados como não aconselháveis são identificados através do nome ou descrição, caso alguma palavra chave contenha termos pejorativos ou de contexto sexual,por exemplo: sex, porn, beijar, transar, flagra sexual e etc, seja encontrada em ambos, o vídeo é classificado como perigoso. Qualquer valor acima de 5 já é considerado como não seguro.
- Conversações: Um campo booleano que recebe “Verdadeiro” ou “Falso” informando se as conversas privadas da criança possui alguma palavra torpe ou conteúdos impróprios.

Por exemplo digamos que tenhamos um objeto Criança, preenchido com os seguintes valores: Postagens: 9.5, Álbuns: Falso, Preferências: Amarelo, Feed de notícias: Falso, Família: Falso, Jogos: 5, Vídeos: 4, Cutucadas: 1 e Conversações: Verdadeiro. Com esse perfil podemos afirmar que os algoritmos de classificação utilizados pelo projeto irão categorizar esse objeto Criança como suscetível a risco, até mesmo por ser um instância já testada pela ferramenta e possuir mais de 5 campos não tão bem catalogados (Postagens, Álbuns, Preferências, Feed de notícias e Família) pelos agentes autônomos. Em virtude de que dos nove campos cinco foram categorizados como não seguros, por heurísticas já definidas pelos agentes, o módulo de mineração de dados informou que essa instância não está sob total segurança, ou seja, se encontra sujeita a riscos.

O sistema possui um agente autônomo que possui por responsabilidade avaliar questões relacionadas aos álbuns das crianças. Detém quatro métodos, são eles:

- Verificador de quantidade de fotos: Se a quantidade de fotos por álbum passar de cem o agente considera o álbum um pouco exposto.
- Verificador de comentários: Checa os comentários maldosos ou aliciadores por foto de cada álbum.

- Verificador de descrição: Confere as descrições em fotos feitas pela criança, caso possua algum chingamento ou palavra torpe, a mesma será armazenada em um local e depois retornada para o relatório final.
- Verificador de álbuns compartilhados: Existem tipos de álbuns, dentre os geralmente mais utilizados para fotos pessoais são os classificados como: normal e móvel (para fotos provenientes do celular), caso esses álbuns sejam do tipo abertos para compartilhamentos, ou seja, amigos ou terceiros possuam permissão para subirem fotos aos mesmos o agente verificador apontará isso como uma brecha de privacidade para a criança.

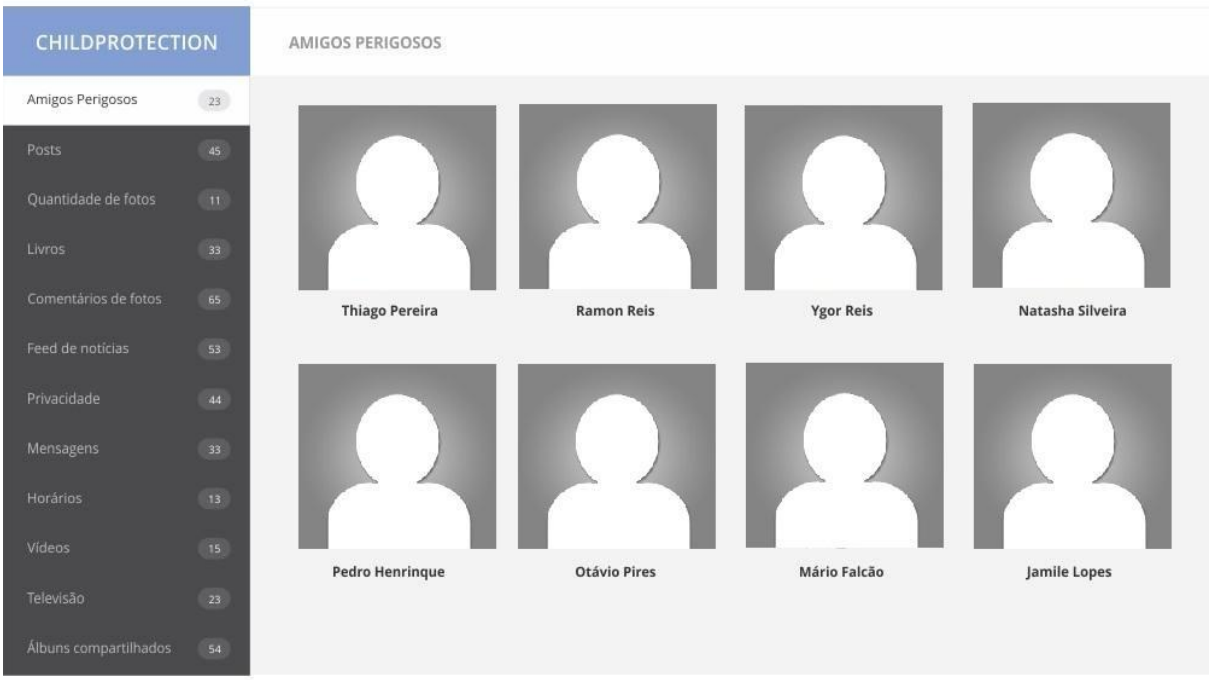
Outro agente autônomo também considerado relevante é o verificador do feed de notícias, o mesmo obtêm as atuais notícias postadas por amigos ou páginas no seu feed e analisa o conteúdo, caso algo “ruim” seja encontrado o nome, foto e mensagem encontrados serão adicionados ao relatório.

Há também um agente responsável por recolher todas as cutucadas durante o período da madrugada (00:00 até 06:00) e adicioná-las ao relatório.

Entre outros agentes podemos citar o analisador de conversas privadas, analisador de conversas externas e o analisador de datas e horários nocivos à criança.

Isto é, se pode deduzir que quem decide o que é certo e o que é errado são os agentes autônomos do sistema e o modulo de mineração treina um modelo que serve para identificar automaticamente se novas ocorrências são de risco ou não para as crianças.

Figura 4 - Página HTML com as informações maliciosas



Fonte: Sistema ChildProtection

A ferramenta de Mineração de dados Weka foi escolhida por ser conhecida como uma das ferramentas “top 10 free” do mercado de Business Intelligence atual, de acordo com o site da “Predicts Analytics Today” (NYCE, 2007). Um dos algoritmos de classificação utilizado é do tipo SVM (Support Vector Machine), também implementado no Weka como SMO (Sequential Minimal Optimization), ambos melhores definidos na seção 4.6. O fluxo principal do processo de Mineração de Dados foi dado da seguinte maneira:

6.2.1 - Seleção dos dados;

6.2.2 - Pré-Processamento;

6.2.3 - Mineração;

6.2.4 - Interpretação e Avaliação dos resultados.

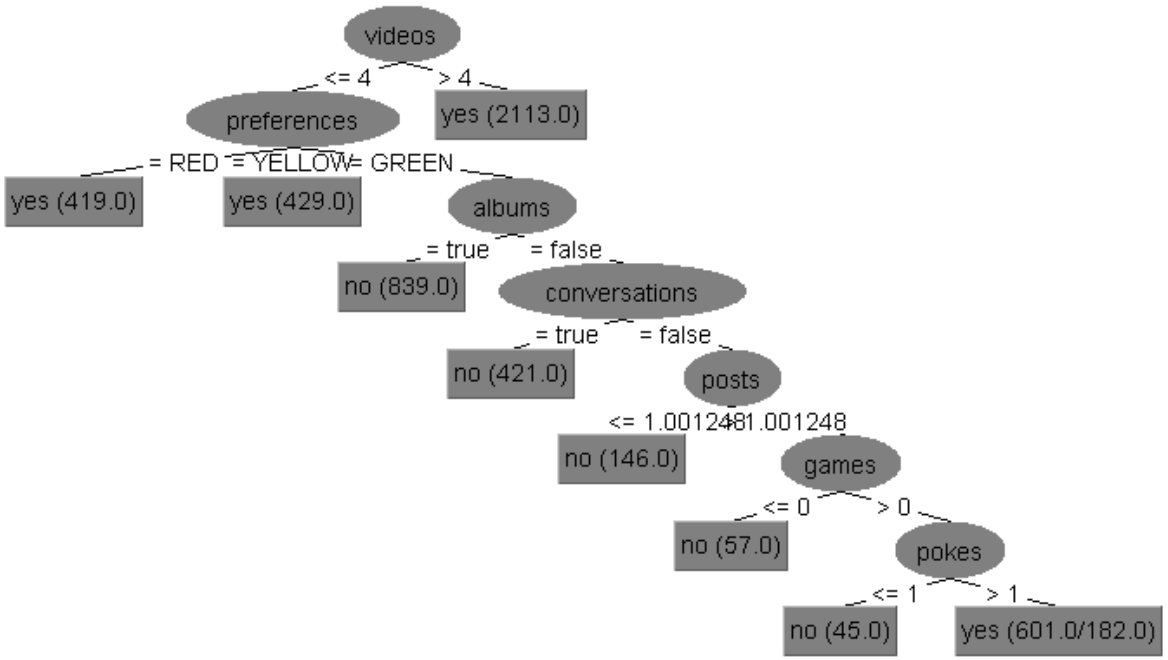
Algumas alterações foram feitas durante a criação do modelo de classificação da criança, melhor descrito na seção 4.5, entre essas podemos citar a adição de dois novos algoritmos de classificação do Weka, o J48 (melhor descrito na seção 4.7) e o IBK (melhor descrito na seção 4.8), ambos foram agregados ao projeto com o intuito de promover maior acurácia na análise de dados e lograr distintas perspectivas feitas pelo Weka. Os três algoritmos utilizados, SMO, J48 e IBK, possuem formas diferentes de aplicar a classificação na mineração de dados, daí a relevância de incorporá-los ao sistema.

O modelo de classificação da criança é formado por colunas com valores, tipos de postagens, álbuns seguros, preferências sobre livros, músicas, vídeos e etc, feed de notícias, família cadastrada na conta ou não, jogos não aconselháveis, vídeos com conotação não infantil, cutucadas na madrugada e conversações com teor adulto, e uma última coluna responsável por classificar a tupla como “há risco” ou “não há risco”. As colunas das tuplas são: postagens, álbuns, preferências, feed de notícias, cutucadas, família, jogos e vídeos. É significativo ressaltar que essas colunas não são todos os dados recuperados para análise (melhor definido na seção 6.1), porém são frutos de análise e seleção afim de reduzir a quantidade de colunas no modelo, deixando apenas aquelas que terão impacto nos resultados.

Inicialmente, o modelo foi gerado baseado nos dados que foram obtidos por meio do uso de heurísticas, definidas na seção 4.4 / 4.5. O modelo obtido tem alta taxa de acurácia e ele é nomeado como modelo de melhor cenário, ou seja, há fraca presença de ocorrências pouco frequentes como falsos positivos e falsos negativos. A princípio possuía uma alta taxa de correção para classificar as instâncias, não por estar correto mas sim por estar sendo um modelo viciado, ou seja, apenas uma coluna com um valor oferecedor de risco o modelo gerado já classificaria essa instância como “há risco” e não se preocuparia em ir mais além nas outras colunas visando gerar um resultado fundamentado no máximo de colunas possíveis.

Todavia após uma sequência de testes o arquivo composto pelos dados responsáveis em gerar o modelo proporcionou um padrão apto para receber diferentes tipos de instâncias, ou seja, não mais fundamentado em apenas uma coluna, como se pode ver na Figura 5, e sim no máximo possível de acordo com os dados da criança a serem recebidos.

Figura 5 - Árvore de decisão modelo classificatório da seção 4.5



Fonte: Ferramenta Weka

A estrutura utilizada para representar o modelo gerado na Figura 5 foi uma árvore de decisão, conceito melhor descrito na seção 4.9, e se pode notar que os atributos estão configurados com valores específicos e, de acordo com as tomadas de decisões, a instância será classificada como “no”, não corre risco, ou “yes” que adverte a criança a um possível perigo.

O Weka possui uma suíte de testes, entre eles existem quatro que foram utilizados nesse trabalho e que serviram de verificação para os dados gerados do modelo, abaixo são eles:

- Use training set: O classificador é avaliado para medir quão bem ele prediz a classe das instâncias do modelo, usando 100% das instâncias (WITTEN, 2005).
- Supplied test set: O classificador é avaliado para medir o quão bem ele prediz a classe de um conjunto de instâncias carregadas a partir de um arquivo específico (WITTEN, 2005).
- Cross-validation: O classificador é avaliado pela validação cruzada, ou seja, usando um certo número pré-definido x de instâncias, todas as instâncias são separadas em “x” partes e vários mini-modelos são construídos cada um com x-1 instâncias, a instância que sobra é usada como teste para aquele mini-modelo e ao final se tem o resultado do modelo geral (WITTEN, 2005).

- Percentage split: O classificador é avaliado o quão bem ele prevê uma certa porcentagem dos dados que se estendeu para o teste. A quantidade de dados mantidos depende do valor inserido no campo % (WITTEN, 2005).

Os testes executados serviram para refinar o modelo e deixá-lo mais apto para categorizar cada instância da forma menos incorreta possível. Em virtude disso a inevitabilidade de se adicionar outliers, identificação de dados que deveriam seguir um padrão esperado mas não o fazem (LIDIO, 2014), se tornou relevante para aprimorar o modelo aos casos menos corriqueiros.

6.2.1 Seleção dos dados

Os dados colhidos foram explorados para haver uma seleção dos mais relevantes para o processo de pré-processamento das informações.

Nessa etapa uma análise cognitiva de como as crianças vem utilizando o Facebook, bem como seus possíveis aliciadores, foi realizada gerando um "modelo" de algoritmo a partir das variáveis definidas na seção 4.4 e assim foi possível implementar a classificação dos suspeitos.

Alguns campos provenientes do Facebook, contexto, atletas favoritos, músicas, jogos, relacionamentos, educação, familiares e dentre outros, foram tratados de acordo com o peso que tiveram na influência para classificação da situação do aliciamento sexual infantil. Para se obter maior precisão no modelo classificador a quantidade de variáveis utilizadas para construção deve ser diversificada a fim de proporcionar diferentes pontos de vista na categorização dos perfis.

6.2.2 Pré-Processamento

A etapa do pré-processamento serviu para fazer a limpeza dos dados buscando evitar redundâncias, elementos ruidosos, incompletos e imprecisos. A limpeza, procedimento realizado manualmente, proporcionou maior acurácia ao modelo classificatório para minerar os dados, por esse motivo se torna viável eliminar tais inconsistências.

Dentro da ferramenta Weka um filtro foi selecionado para fazer o pré-processamento dos dados, o "StringToNominal" serviu para transformar os valores "Verdadeiro" e "Falso" em valores nominais, aqueles atributos que possuem um conjunto de valores pré-definidos,

melhorando a forma como o Weka manipula os dados e atingindo um maior percentual de acerto. Os campos selecionados para serem filtrados com o "StringToNominal" foram: álbuns, preferências, feed de notícias, família e conversações.

6.2.3 Mineração

A mineração propriamente dita foi feita pela ferramenta em perfis de crianças que autorizaram a aplicação, juntamente com a permissão de seus responsáveis, ter conhecimento de suas informações. Dessa forma o grau de risco à exposição, assim como ações comportamentais da criança quanto ao uso da rede social podem ser identificadas sem necessidade de analisar o perfil de terceiros ou ferir seus respectivos direitos de privacidade.

Os algoritmos de classificação utilizados foram o SMO, J48 e o IBK, melhores definidos nas respectivas seções 4.6, 4.7 e 4.8.

6.2.4 Interpretação e Avaliação dos resultados

Por meio do painel de visualização da ferramenta Weka os resultados da mineração dos dados poderão ser obtidos, visualizados, interpretados e avaliados se possuem alguma validade para o problema. A Figura 6, logo abaixo, mostra um exemplo de um resultado:

Figura 6 - Resultado de classificação do algoritmo J48

```
Correctly Classified Instances        4884               96.3314 %
Incorrectly Classified Instances       186                3.6686 %
Kappa statistic                          0.9152
Mean absolute error                      0.0522
Root mean squared error                  0.1634
Relative absolute error                 11.7532 %
Root relative squared error             34.6657 %
Coverage of cases (0.95 level)          99.9606 %
Mean rel. region size (0.95 level)      56.6864 %
Total Number of Instances             5070

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
                 0.999    0.109    0.948      0.999   0.973      0.918  0.992     0.995     yes
                 0.891    0.001    0.999      0.891   0.942      0.918  0.992     0.984     no
Weighted Avg.    0.963    0.073    0.965      0.963   0.963      0.918  0.992     0.992

=== Confusion Matrix ===

    a    b   <-- classified as
 3378    2 |    a = yes
  184 1506 |    b = no
```

Fonte: Ferramenta Weka

Na Figura 6 se pode visualizar o percentual de correção do modelo e também o de incorreção, logo na parte de cima, abaixo se tem também o número total de instâncias, nesse exemplo são 5070, falsos positivos (crianças que correm risco e foram classificadas como seguras) são 2 e os falsos negativos (crianças que não correm risco e foram classificadas como em perigo) são 184.

6.3 Aplicação do sistema no perfil de uma criança

Identificamos e selecionamos um perfil infantil no Facebook e, com permissão dos pais da mesma, utilizamos seus dados para análise. A criança escolhida possui 12 anos, sexo feminino e é natural da cidade de Floriano - PI.

6.4 Avaliação dos resultados obtidos

Após a execução do processo, os resultados obtidos foram informados aos responsáveis e estes determinaram mais precisamente o possível perigo sofrido pela criança. A percepção dos resultados obtidos pela ferramenta foi feita pelos responsáveis da criança avaliada em questão, por intermédio de uma interação com a página HTML, Figura 4, possuinte dos resultados.

Os pais da criança avaliada em questão, melhor definida na seção 6.3, forneceram a autorização para análise já tendo conhecimento que a não divulgação das informações pessoais da criança, como o nome, perfil do Facebook ou qualquer dado que exponha sua dignidade, seriam reveladas. Todavia os resultados poderão ser listados sobre o conteúdo encontrado no perfil da criança. Logo abaixo os dados serão expostos em modo texto.

Datas perigosas: No ano de 2013 a criança usou o Facebook 224 vezes pela madrugada;

Canais de televisão não aconselháveis: Malhação;

Músicas não aconselháveis: Gaby Amarantos, Ivete Sangalo, Banda Garota Safada, Israel Novaes, Lucas Lucco e Claudia leitte;

Família: Nenhum membro da família está cadastrado na sua conta;

Conversações privadas:

Nas conversações privadas algumas palavras não aconselháveis para a idade da criança foram encontradas, como por exemplo: “porra, caralho, fdp, safada e a expressão quero te ver nua dita por um amigo”. Os horários em que esses acontecimentos foram registrados estão em maior parte no período da noite, entre 19:00 horas até 02:00 horas do dia seguinte.

No período do carnaval de 2014 alguns amigos da criança perguntaram se a mesma teria ido para festas, como a de uma banda chamada Chicabana e da cantora Claudia Leitte.

E ao final um amigo em específico da criança a chamou para uma festa em sua casa às 20:24 horas no dia 23 de agosto de 2013.

Cutucadas na madrugada: 22 cutucadas.

Postagens da criança:

“E que se for para ser sozinha e feliz, que minha vida prossiga; D #Happy #Funy #Girl #Top (08/abril/2014 às 12:31:57)”.

Se nota que apesar da conta do Facebook ter ficado inativa no início do ano de 2014 até Junho de 2015 ainda assim alguns dados puderam ser recuperados e identificados como não seguros para uma criança de 12 anos.

7 CONCLUSÕES

Esse trabalho propôs o desenvolvimento de uma ferramenta inteligente que analise os dados provenientes do Facebook de uma criança e informe seu possível nível de vulnerabilidade à riscos. A ferramenta desenvolvida possui características adicionais do que às previamente definidas, como o uso de dois algoritmos de classificação adicionados (IBK e J48), uma interface Web para visualização dos resultados (Figura 4), três novos agentes adicionados (verificador de conversas privadas, verificador de conversas externas e verificador do feed de notícias) e um modelo classificatório mais abrangente para aceitar distintos dados provenientes da conta da criança.

Anteriormente o modelo utilizado para classificação das instâncias provocava divergências nos resultados devido a conflitos nas informações, isto é, instâncias com dados semelhantes categorizadas como classes distintas, após uma sequência de filtragens e escolha correta dos tipos de dados a serem utilizados no modelo, os três algoritmos de classificação começaram a retornar resultados iguais, ou seja, quando o algoritmo SMO classificava uma criança sob risco os outros dois, IBK e J48, também a classificavam igual.

A arquitetura utilizada não gerou problemas, a não ser na parte de sincronização do servidor e o sistema Multiagente, para isso um controle de threads foi necessário para o servidor esperar a resposta dos agentes autônomos e logo em seguida prosseguir com a execução do código.

Antes do sistema não estar plenamente desenvolvido alguns testes foram realizados utilizando-se da ferramenta gráfica de Data Mining Weka na página do Facebook “oficialbarbiebrasil”, essa por sua vez foi uma excelente fonte de dados por possuir uma grande quantidade de curtidas, mais de 12 milhões. Alguns tipos de comentários em fotos publicados pela página possuíam tendências não infantis, informações textuais que fomentavam a persuasão por parte do aliciador. A divulgação das informações não foram exibidas por preservação de ambas as partes envolvidas.

Uma opção plausível para trabalhos futuros seria o teste da ferramenta em novos perfis infantis, com preferência a crianças que possuam atividade assídua no Facebook ou contenham uma quantidade de dados relevantes para análise. E a cada análise verificar novos dados e buscar aprimorar o algoritmo de identificação à palavras torpes com os novos textos analisados, com isso a acurácia do modelo de classificação poderá ser aperfeiçoada.

A forma como os dados são expostos aos responsáveis da criança poderia ser modificada para algo mais automatizado, como por exemplo, a geração de um arquivo PDF

que poderia ser enviado diariamente para o e-mail dos responsáveis relatando o uso da rede social feito pela criança em seus respectivos dias.

O sistema retornou resultados condizentes com a realidade, e possibilitou aos responsáveis da criança, definida na seção 6.3, estabelecerem uma conversação e procurarem uma solução para ajudar a criança a se comportar melhor dentro da rede social e na própria internet de maneira geral, de maneira que a mesma esteja mais protegida contra possíveis aliciadores infantis.

REFERÊNCIAS

ABBOTT, D. W.; MATKOVSKY, I. P.; ELDER, J. F. & IV. An evaluation of highend data mining tools for fraud detection. International Conference on Systems, Man, and Cybernetics, pp. 12--14, IEEE, 2011.

SANTIM, P. L. L., FREITAS, C. O. A.; PARAISO, E. C.Emerson Cabrera Paraiso. Análise automática de textos de mensagens instantâneas para detecção de aliciamento sexual de crianças e adolescentes. V. 2, n. 2, p. 43-59, PUC - Paraná, 2011.

FIRE, D. A. Y. Friend or Foe? Fake Profile Identification in Online Social Networks. Ben gurion. Israel, 2012. Springer Journal of Social Network Analysis and Mining.

BOMBONATTO, Q. - Associação Brasileira de Psicopedagogia. XVI Encontro de Psicopedagogia do Ceará, na UNICHRISTUS. Fortaleza, 2012.

PESSOA, A. S. A., LIMA, G. R. T., SILVA, J. D. S., STEPHANY, S., STRAUSS, C., C, M., FERREIRA, N. J. Meteorological data mining for the prediction of severe convective events, Revista Brasileira de Meteorologia, 2012.

DOS SANTOS, V. S.; BEZERRA, E. P.; ALTURAS, B. Análise de mecanismos de controle de acesso nas redes sociais. Rev. Portuguesa e Brasileira de Gestão, v. 9, n. 3, set. Lisboa, 2010.

SANTOS, Vinicius Souza dos; BEZERRA, Ed Porto; ALTURAS, Bráulio. Análise de mecanismos de controle de acesso nas redes sociais. Rev. Portuguesa e Brasileira de Gestão, Lisboa, v. 9, n. 3, set. 2010 .

GUIMARÃES, S. Brasil é o quarto maior mercado de internet, EXAME, v. 24, n.11, edição 284, p.78-81, São Paulo, 2013.

PEREIRA, S. E. F. N. Redes sociais de adolescentes em contexto de vulnerabilidade social e sua relação com os riscos de envolvimento com o tráfico de drogas. Tese (Doutorado em Psicologia Clínica e Cultura) – Instituto de Psicologia, Universidade de Brasília, Brasília,

2009.

SILVIO, C. 38% das crianças no Facebook têm idade abaixo do permitido, LeiaJá, v. 21, n.7, edição 344, p.18-22, São Paulo, 2012.

MAGNO J., Brasil é o 5º lugar no ranking de cibercrimes, CONGRESSO SERASA EXPIRIAN HITWISE, 1, Porto Alegre, 2014.

CARDOSO O. N. P, MACHADO R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. Revista Adm Pública. 42(3):495-528. 2008.

GOLDSCHIMIDT R, Passos E. Data mining: um guia prático, conceitos, técnicas, ferramentas, orientações e aplicações. Elsevier, São Paulo, 2005.

MARCANO A. YJ, TALAVERA P. R. Minería de datos como soporte a la toma de decisiones empresariales. Opcion. 23(52):104-18. 2007.

JÚNIOR, TARAPANOFF K. Precisão no processo de busca e recuperação da informação: uso da mineração de textos. Ci Inf. 2006;35(3):236-47. 2006.

REUTERS, R. P. Brasil chega a 76 milhões de usuários ativos no facebook, mais da metade acessada pelo celular. São Paulo: Seminare, 2014.

MATSUKI, Como os adolescentes usam internet no Brasil. São Paulo, Portal ebc, 2012.

COOPER, 10 Surprising social media statistics that will make you rethink your strategy. Panama, 2012.

NYCE, Predictive Analytics White Paper, American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America, p. 1, 2007.

KIETZMANN, J.H., HERMKENS, K., McCarthy, I.P., & Silvestre, B.S. Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons,

Vol. 54(3), pp. 241-251. 2011.

LEMIEUX, VINCENT. MATHIEU OUIMET, Sérgio Pereira. Análise Estrutural das Redes Sociais. 2008.

BELLONI, M.L. – O que é Mídia-educação. Campinas, Editora Autores Associados, 2001.

WORTLEY, R. – Pornografia infantil na internet. Washington, Departamento de Justiça dos Estados Unidos, 2012.

FAVERO, D. – Saiba como pedófilos buscam vítimas na internet. São Paulo, ONG/Terre des Hommes, 2014.

BELLIFEMINE, G. C. D. G. Developing multi-agent systems with JADE. [S.l.]: John Wiley & Sons Ltd, 2007. ISBN 978-0-470-05747-6.

CORTES e V. N. Vapnik. Support vector networks. Machine Learning, 20(3) : 273–296, 1995.

WITTEN, Ian H.; Frank, Eibe. Data mining: practical machine learning tools and techniques. San Francisco: Morgan Kaufmann, 2005.

DOMBROWSKI, D. Protecting Children From Online Sexual Predators: Technological, Psychoeducational, and Legal Considerations. Sacramento - California. Professional Psychology: Research and Practice. 2004, Vol. 35, No. 1, 65–7

FINKELHOR, D., MITCHELL, K. J., & WOLAK, .Online victimization: A report on the nation is youth. Alexandria, VA: National Center for Missing and Exploited Children, (2003).

MISHNA, M. W., Family Chair in Child and Family, University of Toronto, Factor-Inwentash Faculty of Social Work, 246 Bloor Street West, Toronto, Ontario M5S 1A1, Canada. 2011.

TUROW, J.The Internet and the family: The view from parents the view from the press. (Report No. 27) .Philadelphia: Annenberg Public Policy Center of the University of Pennsylvania.1999.

HUNTER, Internet Filter Effectiveness—Testing Over- and Underinclusive Blocking Decisions of Four Popular Web Filters. Social Science Computer Review, Vol. 18 No. 2, 2000 214-222, 2000.

GUDJONSSON, G. H., & SIGURDSON, J. F. Differences and similarities between violent offenders and sex offenders.Child Abuse and Neglect, 24,363–372.2000.

ELLIOTT, M., BROWNE, K., & KILCOYNE, J. Child sexual abuse prevention: What offenders tell us.Child Abuse and Neglect, 19,579–594.1995.

MIRANDA, A. O., & CORCORAN, C. L. Comparison of perpetration characteristics between male juvenile and adult sexual offenders: Preliminary results. Sexual Abuse: Journal of Research and Treatment, 12, 179–188.2000.

DAVIS, L., MCSHANE, M. D., & WILLIAMS, F. P. Controlling computer access to pornography: Special conditions for sex offenders.Federal Probation, 59(2), 43–58. 1995.

FIELDMAN, J. P., & CRESPI, T. D. Child sexual abuse: Offenders, disclosure, and school-based initiatives.Adolescence, 37,151–161. 2002.

KENNY, M. C., & MCEACHERN, A. G. Racial, ethnic, and cultural factors of childhood sexual abuse: A selected review of the literature. Clinical Psychology Review, 20,905–922. (2000).

DEREZOTES, D., & SNOWDEN, L. Cultural factors in the intervention of child maltreatment.Child and Adolescent Social Work, 7,161–175. 1990.

GLASSER, M., KOLVIN, I., CAMPBELL, D., GLASSER, A., LEITCH, I., & FARRELLY, S. Cycle of child sexual abuse: Links between being a victim andbecoming a perpetrator.British Journal of Psychiatry, 179,482–494.2001.

U.S. DEPARTMENT OF JUSTICE. Internet crimes against children.Office for Victims of Crime Bulletin. Washington. (2001).

PINHEIRO CAR. Inteligência analítica: mineração de dados e descoberta de conhecimento. Rio de Janeiro: Ed. Ciência Moderna;397 p. 2008.

RYAN H., Dan Fu. Construction a Decision Tree Based on Experience. In: AI Game programming wisdom 2. Charles River Media, 2004.

LIDIO M., Mineração de Dados com Detecção de Outliers em Tarefas de Predição de Séries Temporais, XI Simpósio de excelência em gestão e tecnologia. SEGET. 2014.

WHITLEY, M. POULSEN, S.V. Assertiveness and sexual satisfaction in employed professional women. *Journal of Marriage and Family 37*, 573-581.2006.

BUCKINGHAM, D. After the Death of Childhood: Growing Up in the Age of Electronic Media. Cambridge: Polity Press, 2000.

ALGERIS, SOUZA LM. Violence against children and adolescents: a challenge in the daily work of the nursing team. Rev Latinoam Enferm. 2006.

CHAPELLE, O., SCHÖLKOPF, B., & ZIEN, A. Semi-supervised learning. MIT Press. 2006.

HABIGZANG, K., A., Abuso Sexual Infantil e Dinâmica Familiar:Aspectos Observados em Processos Jurídicos. Psicologia: Teoria e Pesquisa, Vol. 21 n. 3, pp. 341-348, 2005.

AMAZARRY, K., Alguns aspectos observados no desenvolvimento de crianças vítimas de abuso sexual.Psicol. Reflex. Crit. v.11 n.3, Porto Alegre 1998.

SANTIM, P. L. L., FREITAS, C. O. A.; PARAISO, E. C.Emerson Cabrera Paraiso. Análise automática de textos de mensagens instantâneas para detecção de aliciamento sexual de crianças e adolescentes. V. 2, n. 2, p. 43-59, PUC - Paraná, 2011.