



# **Identificação de perfis falsos nas redes sociais**

Mestrado em Cibersegurança e Informática Forense

Hugo Filipe Fontainhas Baptista

Leiria, Setembro de 2019



# **Identificação de perfis falsos nas redes sociais**

Mestrado em Cibersegurança e Informática Forense

Hugo Filipe Fontainhas Baptista

Projeto realizado sob a orientação do Professor Doutor Mário João Gonçalves Antunes

Leiria, Setembro de 2019

# **Originalidade e Direitos de Autor**

O presente relatório de projeto é original, elaborado unicamente para este fim, tendo sido devidamente citados todos os autores cujos estudos e publicações contribuíram para a/o elaborar.

Reproduções parciais deste documento serão autorizadas na condição de que seja mencionado o Autor e feita referência ao ciclo de estudos no âmbito do qual o mesmo foi realizado, a saber, Curso de Mestrado em Cibersegurança e Informática Forense, no ano letivo 2018/2019, da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria, Portugal, e, bem assim, à data das provas públicas que visaram a avaliação destes trabalhos.

# Agradecimentos

Um projeto é um trabalho individual, mas a qual conta com o apoio, ajuda, esforço de um elevado e amplo leque de pessoas, que de uma forma ou de outra, deram um pouco de si ou do seu tempo para que este projeto se concretizasse.

Em primeiro lugar quero agradecer ao Professor Baltazar Rodrigues e ao Professor Doutor Mário João Gonçalves Antunes, por terem acompanhando e orientado este projeto desde início, dando uma preciosa visão sobre o tema. Agradeço especialmente pela ajuda com as observações perspicazes e pelo incentivo constante.

Em segundo lugar agradeço a todas aquelas pessoas que ficaram ao meu lado durante todo este tempo incentivando-me, dando-me força, mesmo quando eram privadas da minha companhia e abdicavam um pouco da sua vida. Aquelas pessoas que acreditaram sempre, mesmo quando eu não acreditava.

Aos meus colegas de luta diária, sempre presentes, João Santos e Nuno Rasteiro, sem esta força conjunta e apoio possivelmente este projeto não teria este final.

# Resumo

Atualmente com o constante crescimento das redes sociais e com o papel que desempenham na sociedade, tanto a nível social como de negócios, torna-se importante compreender alguns problemas que se têm vindo a identificar nessas redes. Para os seus utilizadores a sua vida prática do dia-a-dia, tornou-se interligada pelas redes sociais.

A popularidade das redes sociais acarreta alguns problemas, como por exemplo a possibilidade de expor informações pessoais dos utilizadores, a propagação do spam, a exposição a cenários de potencial extorsão e outras atividades relacionadas com o cibercrime.

Uma das formas de potenciar estes episódios de cibercrime está relacionado com a criação de perfis falsos e com a sua utilização em atividades ilícitas e potencialmente lesivas para os utilizadores das redes sociais.

No âmbito deste projeto foi desenvolvido um estudo sobre o funcionamento das redes sociais e dos principais crimes praticados com recurso a perfis falsos, bem como os principais mecanismos existentes para identificar esse tipo de perfis.

Este projeto teve como principal objetivo o desenvolvimento de uma metodologia para auxiliar na identificação de perfis falsos na rede social Twitter. Para tal foram identificados vários parâmetros relacionados com a conta dos utilizadores nessa rede social e, para cada parâmetro, foi calculado o seu valor ponderado, com base na informação pública do perfil em apreciação.

O resultado obtido da análise de cada perfil permite aferir sobre a probabilidade de o mesmo ser falso, ou não. No âmbito deste projeto foi desenvolvida uma aplicação web que implementa a metodologia definida e calcula a probabilidade de “falsidade” do perfil. A disponibilidade alargada da aplicação permite a consulta em tempo real e de forma rápida, em cada momento, do nível de “falsidade” de uma determinada conta.

Os testes foram realizados com dois *datasets* conhecidos e já publicados, correspondendo a conjuntos de perfis falsos e verdadeiros no Twitter. Além de se ter validado a viabilidade da solução e aplicação da metodologia na rede social Twitter, foi possível obter resultados promissores através da elevada assertividade, tendo-se registado uma percentagem média de acerto de 94.5%.

**Palavras-chave:** Redes sociais, perfis falsos, metodologia, Twitter.

# Abstract

Nowadays, with the consistent growth of social networks and with the role they play, both socially and in business, it is important to understand some problems that have been identified in these networks. For their users, the day to day life has become interconnected with these social networks.

The fame of social networks has brought some problems, such as the possibility of exposing users' personal information, the spread of spam, exposure to potentially extortionate and others cybercrime-related activities.

One of the ways to boost cybercrime episodes is related to fake profile creation on and whit its use on illicit activities and potentially harmful for the users of such social networks.

It has been developed, with this work a study of social networks, crime committed, using fake profiles, existing studies and mechanisms to identify these fake profiles.

The main goal was to develop a methodology that helps to identify fake profiles. For this purpose, a set of parameters, based on public information of an existing profile on social network, was used to verify the veracity of the profile.

The result obtained from the analysis of each profile allows us to assess the probability that it is false, or not. Another goal was to develop an application, that allows to execute this research, based on the proposed methodology, and return a percentage of falsity of an profile. The wide availability of the application allows quickly and in real time to make a search, at any moment, and to know the “fakeness” of a profile.

In addition, the application texts made whit two well-known and published datasets, corresponding to fake and true profile sets on Twitter. In addition to validating the viability of the solution, and applying the methodology on the Twitter social network, promising results were achieved through high assertiveness, with an average hit percentage of 94.5%.

**Keywords:** social networks, fake profiles, methodology, Twitter.



# Índice

<b>Originalidade e Direitos de Autor</b> .....	<b>iii</b>
<b>Agradecimentos</b> .....	<b>iv</b>
<b>Resumo</b> .....	<b>v</b>
<b>Abstract</b> .....	<b>vi</b>
<b>Índice</b> .....	<b>viii</b>
<b>Lista de Figuras</b> .....	<b>x</b>
<b>Lista de tabelas</b> .....	<b>xi</b>
<b>Lista de siglas e acrónimos</b> .....	<b>xii</b>
<b>1. Introdução</b> .....	<b>1</b>
<b>1.1. Contexto</b> .....	<b>1</b>
<b>1.2. Motivação</b> .....	<b>2</b>
<b>1.3. Objetivos</b> .....	<b>3</b>
<b>1.4. Estrutura do Documento</b> .....	<b>3</b>
<b>2. Conceitos Fundamentais</b> .....	<b>5</b>
<b>2.1. Redes sociais</b> .....	<b>5</b>
2.1.1. Definição .....	5
2.1.2. Tipos de redes sociais.....	6
2.1.3. Facebook .....	6
2.1.4. Instagram.....	8
2.1.5. Twitter .....	8
<b>2.2. APIs</b> .....	<b>9</b>
2.2.1. Twitter API.....	9
2.2.2. Facebook API.....	11
2.2.3. Instagram API .....	12
2.2.4. Limitações .....	13
<b>2.3. Noção de Perfil Falso</b> .....	<b>14</b>
2.3.1. Perfis Falsos nas Redes Sociais.....	15
2.3.2. Crime com recurso às redes sociais.....	16
<b>2.4. Estado da Arte - Detecção de Perfis Falsos nas Redes Sociais</b> .....	<b>18</b>
<b>3. Arquitetura</b> .....	<b>25</b>



<b>3.1.</b>	<b>Abordagem .....</b>	<b>25</b>
<b>3.2.</b>	<b>Desenho da Arquitetura .....</b>	<b>27</b>
<b>3.3.</b>	<b>Considerações sobre o funcionamento.....</b>	<b>30</b>
<b>3.4.</b>	<b>Desenvolvimento .....</b>	<b>31</b>
<b>4.</b>	<b>Testes e Análise de Resultados .....</b>	<b>35</b>
<b>4.1.</b>	<b><i>Dataset</i> de testes para a rede social Twitter .....</b>	<b>35</b>
4.1.1.	Origem .....	35
4.1.2.	Criação do <i>dataset</i> .....	35
<b>4.2.</b>	<b>Definição de testes e metodologia .....</b>	<b>36</b>
<b>4.3.</b>	<b>Resultados obtidos .....</b>	<b>36</b>
<b>4.4.</b>	<b>Análise de resultados .....</b>	<b>37</b>
<b>5.</b>	<b>Conclusões e Trabalho Futuro .....</b>	<b>41</b>
	<b>Bibliografia ou Referências Bibliográficas.....</b>	<b>42</b>
	<b>Anexos.....</b>	<b>1</b>

# Lista de Figuras

Figura 1 - Exemplo de um pedido à API do Twitter.....	9
Figura 2 - Exemplo de uma resposta da API do Twitter.....	10
Figura 3 - Exemplo de um pedido de um <i>token</i> à API do Facebook. ....	11
Figura 4 - Exemplo de um pedido à API do Facebook.....	11
Figura 5 - Modelo de resposta da API do Facebook do nome e id de um utilizador.....	12
Figura 6 - Conjunto de parâmetros proposto por diferentes investigadores [54].....	20
Figura 7- Conjunto de 22 parâmetros utilizado na investigação. ....	21
Figura 8 - Fórmula da Precision [70].....	22
Figura 9 - Fórmula da Recall [70] . ....	22
Figura 10 - Classificação dos 10 parâmetros e seu peso.....	23
Figura 11 - Classificação dos 7 parâmetros com ponderação igual ou superior a 50%.....	23
Figura 12 - Esquema de Alto Nível de funcionamento da solução. ....	29
Figura 13 - Esquema de funcionamento detalhado da solução. ....	30
Figura 14 - Exemplo de uma página de perfil de um utilizador do Twitter.....	32
Figura 15 - Exemplo de utilização da aplicação.....	33
Figura 16 - Exemplo da página do resultado de uma pesquisa de um perfil.....	33
Figura 17 - Página com informação do projeto e detalhes dos parâmetros pesquisados. ....	34
Figura 18 - Média percentagem falsidade.....	38

# Lista de tabelas

Tabela 1 - Redes Sociais existentes em 2019. ....	6
Tabela 2 - Informação dos <i>dataset</i> do projeto MIB. ....	35
Tabela 3 - Média da percentagem de falsidade.....	36
Tabela 4 - Número de contas por percentagem de falsidade, <i>dataset</i> contas reais. ....	36
Tabela 5 - Número de contas por % de falsidade, <i>dataset</i> contas falsas. ....	37
Tabela 6 - Falsos negativos, <i>dataset</i> contas reais, conjunto 7 parâmetros. ....	39
Tabela 7 - Percentagem Assertividade, <i>dataset</i> contas reais, conjunto 7 parâmetros. ....	39
Tabela 8 - Falsos positivo, <i>dataset</i> contas falsas, conjunto 11 parâmetros. ....	40
Tabela 9 - Percentagem Assertividade, <i>dataset</i> contas falsas, conjunto 11 parâmetros.....	40

## Lista de siglas e acrónimos

ESTG	Escola Superior de Tecnologia e Gestão
API	Application Programming Interface
CAPTCHA	Completely Automated Public Turing Test to Tell Computers and Humans Apart
CNE	Comissão Nacional Eleições
CSV	Comma-Separeted Value
EFFP	Entidade Fiscalizadora do Financiamento dos Partidos
FN	False Negative
FP	False Positive
IP	Endereço IP
JSON	Javascript Object Notation
MIB	My Information Bubble Project
ONU	Organização das Nações Unidas
OSINT	Open Source INTelligence
REST	Representational State Transfer
RGPD	Regulamento Geral de Proteção de Dados
RSS	Really Simple Syndication
SMS	Short Message System
TN	True Negative
TP	True Positive
URL	Uniform Resource Locator

VPN	Virtual Private Network
XLSX	Microsoft Excel Open XML Spreadsheet



# 1. Introdução

Neste capítulo definem-se o contexto e a motivação do trabalho descrito, assim como a estrutura do documento.

## 1.1. Contexto

O constante crescimento das redes sociais e o papel que desempenham na sociedade de hoje em dia, tanto a nível social como de negócios, implica que se torne importante estudar aprofundadamente o seu funcionamento, uma vez que, para os seus utilizadores, a partilha das suas atividades do dia-a-dia pode ter um efeito negativo e por vezes, atentatório da sua liberdade.

A popularidade das redes sociais tem acarretado alguns problemas, como por exemplo a possibilidade de expor informações confidenciais dos utilizadores, a propagação do spam, a exposição a cenários de potencial extorsão e outras atividades relacionadas com o cibercrime.

Uma das formas de potenciar estes episódios de cibercrime está relacionado com a criação de perfis falsos. São vários os casos reportados em que a informação veiculada através de perfis falsos nas redes sociais tem causado enormes danos no mundo real para a sociedade em geral e entidades empresariais entre outros.

O constante aumento do crime com recurso às redes sociais têm-se sentido cada vez com mais incidência, de tal forma que diariamente aparecem notícias demonstrando esse facto, onde os autores passam normalmente impunes devido à criação de perfis falsos [1][2][3][4].

Os perfis falsos são criados com o objetivo de anonimizar o proprietário da conta ou com a alinação de identidade, tornando-se por isso num desafio para as forças policiais, na medida em que se tornam difíceis de identificar e rastrear.

O impacto negativo que os perfis falsos trazem às redes sociais e à sociedade, fizeram com que as próprias redes sociais tentassem combater este sufrágio, implementando mecanismos de segurança no registo, como por exemplo o uso “*captchas*” ou a validação de email, solicitando o número telemóvel para posterior envio de código de validação, etc.

Também a sociedade está atenta ao problema e já vários investigadores tentaram arranjar uma solução para o problema, apresentando várias abordagens como o uso de mecanismos de segurança, recurso a inteligência artificial ou o uso de algoritmos de classificação.

O principal desafio está em entender como são criados e como atuam os perfis falsos, com vista a apresentar uma solução que, com elevado nível de fiabilidade, consiga ajudar a identificar um perfil falso e que seja uma boa base para ajudar e alertar a sociedade que um determinado perfil possa ser falso.

Este trabalho é o resultado de um estudo realizado no âmbito desta temática, que teve como principal objetivo a apresentação de uma solução que possa identificar um perfil falso, com elevado nível de assertividade. Pretende-se desta forma contribuir para que a sociedade esteja mais informada sobre a existência e utilização de perfis falsos nas redes sociais e tenha uma forma de ser auxiliada na tomada de decisão.

O desenvolvimento foi realizado na rede social Twitter. Os testes foram realizados com base em datasets já publicados, com perfis da rede social Twitter. Os resultados obtidos relevaram-se promissores, tendo-se obtido elevadas taxas de acerto.

## **1.2.Motivação**

Os utilizadores das redes sociais, como o Twitter, necessitam de ter ferramentas ágeis que permitam, de forma rápida e imediata, validar se um determinado perfil é fidedigno ou falso. Desta forma, tendo em conta o resultado obtido pelos algoritmos implementados nestas ferramentas, normalmente sob a forma de aplicação, é possível identificar perfis potencialmente falso e, por isso, suscetíveis de divulgar informação falsa (por exemplo associados às *fake news*) ou estarem associados a práticas ilícitas de cibercrime.

Tal é verdade para perfis relativos a utilizadores comuns, como também para empresas ou organizações, que possam vir a ser afetadas por alguém menos bem-intencionado.

Com a realização deste projeto desenvolveu-se uma metodologia e um algoritmo para apoiar a deteção de perfis falsos. Através de uma aplicação web criada especificamente para o efeito, foi possível alargar a utilização da metodologia e do algoritmo a todos os utilizadores. A sua interface web simples potêcia a utilização mesmo por utilizadores com pouca experiência.

Além dos utilizadores comuns, a investigação pelos órgãos de polícia criminal (OPC) poderá igualmente beneficiar da utilização destas aplicações de análise automática, com vista à identificação de perfis associados a atividade criminosa. É por isso também motivação para este trabalho o contributo que pode ter, embora indiretamente, na diminuição do cibercrime nas redes sociais, em especial no Twitter.

Esta metodologia e o algoritmo definido poderá igualmente ser alargado a outras redes sociais, desde que sejam ultrapassadas as limitações no acesso à utilização da API.

O trabalho de investigação já desenvolvido nesta área, nomeadamente ao nível da identificação dos parâmetros de avaliação dos perfis e também dos datasets já publicados, constitui igualmente uma mais-valia, tendo permitido adicionar valor ao trabalho já existente e também motivar o desenvolvimento de futuras aplicações, metodologias e algoritmos para deteção automática de perfis falsos.



### **1.3.Objetivos**

Tendo em conta o problema de identificação de perfis falsos bem como o seu crescimento nas redes sociais e o impacto que isso tem na vida das pessoas e empresas, assim como nas próprias redes sociais, tornou-se necessário tentar resolver esse problema.

Para facilitar a identificação desses perfis falsos definiu-se como objetivo principal desenvolver uma metodologia para ajudar na identificação dos mesmo. Para tal, através da aplicação desenvolvida para o efeito, o utilizador poderá pesquisar um perfil de uma rede social, recorrendo à informação que esta disponibiliza, obtendo como resultado uma probabilidade de o mesmo ser falso ou fidedigno.

Os resultados são calculados com base na análise de vários parâmetros já publicados em [5]. Ao trabalho já desenvolvido foram adicionados mais parâmetros que nos pareceram relevante ao longo do projeto, tendo sido feita uma comparação do desempenho dos vários grupos de parâmetros identificados.

O presente trabalho incide sobre as redes sociais Facebook, Instagram e Twitter, sobre as quais é realizado um enquadramento e uma descrição do seu funcionamento.

A definição da metodologia e do algoritmo desenhados para este projeto inicia-se pelo estudo dos parâmetros associados ao perfil público de um utilizador no Twitter, que permitam garantir uma elevada assertividade no cálculo do índice de falsidade de um perfil.

Será de seguida realizado um estudo das API das redes sociais em análise e dos seus métodos, de modo a que possa automatizar da melhor forma o cálculo dos resultados, com destaque para o funcionamento da API do Twitter.

Como resultados principais deste trabalho destacam-se o relatório com a descrição das decisões tomadas, a metodologia e algoritmo para a operacionalização da validação de um perfil como sendo falso ou fidedigno e uma aplicação web para facilitar o acesso aos resultados obtidos.

Os resultados obtidos revelaram-se promissores quer em desempenho quer em grau de assertividade dos resultados obtidos. Os datasets utilizados para os testes, já publicados anteriormente, permitiram ter boas indicações sobre o desempenho do algoritmo, designadamente ao nível do número de falsos positivos, falsos negativos, verdadeiros positivos e verdadeiros negativos.

### **1.4.Estrutura do Documento**

Este relatório está dividido em cinco capítulos de forma a facilitar a leitura e entendimento do tema em estudo.

Após a introdução, contextualização, motivação e objetivos apresentados no presente capítulo, o capítulo 2 apresenta os conceitos fundamentais sobre as redes sociais, designadamente a sua história e alguns dados relevantes sobre estas, as API e exemplos de

um pedido, assim como as limitações existentes nas mesmas. Neste capítulo será introduzida a uma definição de “perfil falso”, no contexto deste projeto. A apresentação de alguns crimes relacionados com o uso de perfis falsos e o estado da arte em relação à detecção de perfis falsos termina este capítulo.

No capítulo 3 apresenta-se a arquitetura da solução apresentada, as decisões tomadas e a metodologia definida para a detecção do nível de falsidade ou fidedignidade de um perfil. Detalham-se ainda os parâmetros dos perfis dos utilizadores que são utilizados para a aplicação na rede social Twitter. Este capítulo inclui ainda um desenho da arquitetura e do *workflow*, assim como algumas considerações que se deverão ter em conta no funcionamento da solução, informações sobre o desenvolvimento e uma breve explicação sobre a forma de utilização da aplicação.

O capítulo 4 descreve a criação e a origem dos datasets utilizados para a realização dos testes de desempenho, bem como a definição dos testes e a metodologia definida para a elaboração dos mesmos, os resultados obtidos e a correspondente análise.

O capítulo 5, apresentam-se as conclusões sobre o trabalho realizado e apresentam-se algumas linhas de desenvolvimento interessantes num trabalho futuro.

No final do documento encontra-se a bibliografia consultada para a realização deste projeto, assim como os anexos, constituídos essencialmente pelas tabelas detalhadas, com os resultados obtidos.

## 2. Conceitos Fundamentais

Neste capítulo definem-se os conceitos fundamentais para a compreensão dos capítulos subsequentes do documento, nomeadamente a definição de rede social, tipos de redes sociais existentes, descrição das redes sociais Facebook, Instagram e Twitter.

### 2.1. Redes sociais

Nesta secção, além do enquadramento geral do conceito de rede social, é apresentado o conceito de API e a descrição das API do Facebook, Instagram e Twitter. Por fim, é apresentado um exemplo de funcionamento de um pedido e da sua resposta, através de cada uma das API abordadas.

É avançada uma noção de perfil falso, de acordo com o enquadramento deste projeto e são apresentadas as principais ameaças relacionadas com este tipo de perfis nas redes sociais.

#### 2.1.1. Definição

O termo “*rede social*” pode definir-se como um espaço virtual na Internet onde os utilizadores (pessoas, entidades ou empresa) podem criar ligações entre si, com o intuito de partilhar informações.

As redes sociais transmitem a sensação de proximidade e relacionamento virtual. Visam aproximar famílias, reunir pessoas separadas pela distância física num ambiente virtual, geram amores e juntam pessoas com os mesmos interesses ou crenças. Enfim, transmitem a sensação de que qualquer pessoa que esteja numa dessas redes sociais está à distância de um *click*.

A utilização eficiente e correta das redes sociais poderá traduzir-se em excelentes ferramentas de marketing ao serviço das empresas, que a qualquer momento conseguem lançar uma campanha publicitária chegando eficazmente aos seus seguidores em qualquer parte do mundo.

Hoje em dia existem dezenas de redes sociais na internet, segundo Danah M. Boyd e Nicole B. Ellison a primeira rede social que surgiu na internet foi em 1997 [6], SixDegrees.com que permitia aos utilizadores criar perfis e listar os seus amigos, a partir daí foram surgindo muitas outras até aos dias de hoje.

É difícil ter noção da quantidade exata de redes sociais existentes neste momento, pois é algo muito dinâmico, algumas aparecem frequentemente e outras que terminam. A título de curiosidade e para se ter a noção da quantidade de redes sociais existentes, apresento na Tabela 1, uma lista de 65 redes sociais existentes em 2019, [7]:

**Tabela 1 - Redes Sociais existentes em 2019.**

Facebook	LinkedIn	Care2	YY	We Heart It
WhatsApp	Telegram	CafeMom	Snapfish	Buzznet
QQ	Reddit	Ravelry	Friendster	DeviantArt
WeChat	Taringa	Nextdoor	Vkontakte (VK)	Flickr
QZone	Foursquare	Wayn	ReverbNation	MeetMe
Tumblr	Renren	Cellufun	Funny or Die	Meetup
Instagram	Tagged	YouTube	Pinterest	Tout
Twitter	Badoo	Vine	Gaia Online	Mixi
Baidu Tieba	Myspace	Classmates	Snapchat	Douban
Skype	StumbleUpon	MyHeritage	Delicious	Vero
Viber	The Dots	Viadeo	LiveJournal	Quora
Sina Weibo	Kiwibox	Xing	TikTok	Spreely
LINE	Skyrock	Xanga	Discord	

### 2.1.2. Tipos de redes sociais

Há vários tipos de redes sociais, destacando-se as relacionadas com relacionamentos, entretenimento, profissionais e temáticas:

- As redes sociais de relacionamento têm como principal objetivo criar ligações entre as pessoas, a partilha de conteúdos e informações entre elas, como exemplo temos o Facebook.
- Redes sociais de entretenimento como como principal objetivo disponibilizar conteúdos para os utilizadores, como exemplo o Youtube e Pinterest.
- No que diz respeito a redes sociais profissionais, o seu objetivo é interligar utilizadores criando vínculos profissionais entre estes, como exemplo o LinkedIn, onde é possível as empresas registarem-se, colocarem vagas para empregos ou interligarem-se com os seus funcionários.
- Redes temáticas são as que atendem a um nicho de mercado tendo um público alvo muito específico, como por exemplo o Tripadvisor cujo setor alvo é o turismo e a gastronomia.

Descrevem-se de seguida três redes sociais, Facebook, Twitter e Instagram, que pela sua dimensão e relevância merecem uma análise mais cuidada. O desenvolvimento deste trabalho foi baseado no funcionamento da rede social Twitter.

### 2.1.3. Facebook

O Facebook começou em Janeiro de 2004 com o nome de “the facebook” [8]. Este website baseava-se num que já tinha sido criado anteriormente pelo estudante da Universidade de

Harvard, Mark Zuckerberg, e três amigos – Eduardo Saverin, Chris Hughes e Dustin Moskovitz, com o nome de “facemash”[9].

O facemash, criado em 28 Outubro de 2003, tinha como objetivo ser um jogo em que os utilizadores, neste caso os estudantes da universidade, selecionariam entre duas fotos de dois alunos qual a que achariam mais atraente: “Hot or Not”. Logo nas primeiras 4 horas de funcionamento o facemash conseguiu cerca de 450 utilizadores e 22000 visualizações de fotos.

O website acabou por ter uma vida curta devido a uma ilegalidade cometida por Mark Zuckerberg. De facto, para conseguir as fotos de todos os estudantes, ele entrou na rede de Harvard e roubou todas essas fotos, tendo sido expulso da universidade por esse motivo.

Mais tarde em Fevereiro de 2004 foi lançado o site “thefacebook”, no endereço “thefacebook.com”. Tinha acesso restrito apenas aos estudantes de Harvard e após o primeiro mês mais de metade dos estudantes da universidade já se tinham registado. Em março de 2004 o site expandiu-se para as universidades de Stanford, Columbia, e Yale, tendo esta expansão continuado à maioria das universidades dos Estados Unidos e Canadá [10][11].

Em 2005 o “the” foi retirado do nome e ficou apenas “Facebook”. Em outubro desse mesmo ano o site já chegava a 21 universidades do Reino Unido, assim como universidades do México e Porto Rico [12]. No final de 2005 o Facebook já contava com cerca de 2000 universidades, tendo em 26 Setembro de 2006 sido aberto para todas as pessoas que tivessem mais de 13 anos e um endereço de e-mail [6].

Atualmente o Facebook é a rede com mais utilizadores do mundo e as suas receitas provêm maioritariamente de publicidade e grupos patrocinados. Qualquer pessoa pode registar-se e criar um perfil, adicionar outros utilizadores à sua rede de amigos e partilhar fotos, vídeos, mensagens, interesses, criar grupos ou pertencer a grupos assim como criar ou aderir a eventos.

A informação de um utilizador ou grupo pode ser restrita ou pública. Esta definição é gerida pelo utilizador ou grupo, existem também outras funcionalidades como um mural que é um espaço onde os amigos podem escrever mensagens para o utilizador ler. Este mural é visível a todos os utilizadores que possam ver o perfil completo. Existem também aplicações com várias funcionalidades tais como jogos e eventos.

O Facebook disponibiliza também a aplicação para dispositivos móveis, quer sejam Android ou IOS.

No terceiro trimestre de 2018 o Facebook possuía cerca 2.23 mil milhões de utilizadores ativos mensalmente, sendo 1,74 mil milhões em dispositivos móveis, dos quais 1.57 mil milhões de utilizadores ativos por dia via dispositivos móveis e cerca de 1.47 mil milhões em computadores fixos [13].

#### 2.1.4. Instagram

O Instagram é uma rede social online que permite a partilha com outros utilizadores de fotos e pequenos vídeos, designados “Insta Stories”, que só ficam disponíveis durante 24 horas. Um dos recursos que esta aplicação possui é a capacidade de aplicar filtros às fotos, o que motivou o seu elevado reconhecimento e o sucesso que atingiu.

Esta rede foi criada em Outubro de 2010 por Kevin Systrom e Mike Krieger tendo primeiramente apenas sido disponibilizada para aparelhos Apple. em 26 Fevereiro de 2013 atingiu os 100 milhões de utilizadores [14].

Em Abril de 2012 o Facebook comprou o Instagram por cerca de mil milhões de dólares [15], tendo em Setembro de 2018 atingido o patamar de mil milhões de utilizadores ativos por mês, sendo que destes, 500 mil utilizadores estão ativos diariamente [16].

Desde o lançamento até Setembro 2018 já foram adicionadas cerca de 50 mil milhões de fotos, no mês referido anteriormente foram publicadas cerca de 400 milhões de *stories* por dia, 4,2 milhões de *likes* e mais de 100 milhões de fotos e vídeos [16].

#### 2.1.5. Twitter

O Twitter é uma rede social e *blogging* que permite a troca de mensagens entre utilizadores. As mensagens, designadas por *tweets*, são curtas e podem ir até aos 280 caracteres.

Os *tweets* podem ser enviados através do próprio website ([www.twitter.com](http://www.twitter.com)), por SMS ou por meio de aplicações específicas, como por exemplo “Fenix 2”[17], “Friendly for twitter”[18] . Os *tweets* são publicados em tempo real e os seguidores podem visualizá-los na sua conta no próprio website, por SMS e/ou por RSS.

O Twitter foi criado em Março de 2006 por Jack Dorsey, Evan Williams, Biz Stone e Noah Glass, mas só foi lançado uns meses mais tarde em julho de 2016 nos EUA. A ideia inicial pretendia que o Twitter fosse uma espécie de SMS da Internet. Na altura os *tweets* eram apenas de 140 caracteres tendo depois em 26 de Setembro de 2017 sido alteradas para 280 [19].

Em Setembro de 2018 o Twitter possuía cerca de 335 milhões de utilizadores ativos por mês e eram enviados cerca de 500 milhões de *tweets* por dia. A maioria dos utilizadores, cerca de 80%, utiliza dispositivos móveis e o número de utilizadores ativos por dia é cerca de 100 milhões [20].

Segundo um estudo da ONU a população mundial em 2017 era de 7.6 mil milhões de pessoas. Se somarmos o número de utilizadores ativos por mês das três redes sociais, obtemos cerca de 3.56 mil milhões de utilizadores, ou seja, cerca de metade da população mundial estaria ativa numa destas redes sociais [21].

## 2.2.APIs

As Application Programming Interface (API) fornecem uma forma das aplicações comunicarem entre si sem terem de conhecer, por exemplo, a estrutura da BD ou linguagem de programação presente em cada uma. Para tal utilizam funções e padrões (protocolos) de comunicação definidos por cada API.

Nas secções seguintes serão descritas as API das redes sociais mais utilizadas, designadamente o Facebook, Twitter e Instagram.

### 2.2.1. Twitter API

A rede social Twitter não dispõe de apenas uma API mas sim de várias, complementadas por um conjunto de aplicações. A API *standard* é a única totalmente grátis, estando a *Premium* e a *Enterprise* sujeitas a pagamento. No entanto estas apresentam um maior nível de acesso à informação e maior crescimento [22].

A API *standard* tem num conjunto de REST API e *Streaming* API. A *Enterprise* inclui um conjunto de filtros, permite pesquisa histórica, inclui outras API para uma pesquisa de dados mais profunda assim como mecanismo de escuta e disponibiliza aplicações específicas para empresas, sendo cada uma delas seja paga individualmente. Detém ainda o nível máximo de acesso e disponibilidade, assim como gestão de conta e suporte técnico.

A versão *Premium* é taxada consoante o que se utiliza e é indicada para quem a API *standard* já não é suficiente.

Para poder utilizar qualquer API do Twitter é necessário primeiramente estar registado como *developer* e ter criado um conjunto de credenciais e *token*.

O pedido à API da Figura 1 é um pedido HTTP do tipo “GET” cuja resposta virá em formato JSON, representa o pedido dos membros de uma lista.

```
GET
https://api.twitter.com/1.1/lists/members.json?slug=team&owner_screen_
name=twitterapi&cursor=-1
```

Figura 1 - Exemplo de um pedido à API do Twitter.

A Figura 2 apresenta o exemplo da resposta ao pedido anterior.

```
{ "previous_cursor": 0, "previous_cursor_str": "0", "next_cursor": 0,
"users": [ { "profile_sidebar_fill_color": "bedcfa", "expanded_url":
null, "profile_sidebar_border_color": "85add6", "name": "Sharon Ly",
"profile_background_tile": false, "location": "", "profile_image_url":
"http://a2.twimg.com/profile_images/1359867172/image_normal.jpg",
"created_at": "Sun May 25 00:29:44 +0000 2008", "follow_request_sent":
null, "is_translator": false, "profile_link_color": "955ea6", "id_str":
"14895163", "entities": { "urls": [ ], "hashtags": [ ], "user_mentions":
[ ] }, "default_profile": false, "favourites_count": 63,
"contributors_enabled": false, "url": null, "id": 14895163,
"utc_offset": -28800, "profile_image_url_https":
"https://si0.twimg.com/profile_images/1359867172/image_normal.jpg",
"profile_use_background_image": true, "listed_count": 43, "lang": "en",
"profile_text_color": "4c58a3", "followers_count": 784, "protected":
false, "profile_background_color": "312040", "geo_enabled": true,
"description": "", "time_zone": "Pacific Time (US & Canada)",
"verified": false, "profile_background_image_url_https":
"https://si0.twimg.com/profile_background_images/257176598/hydrangeas_
94_68830.jpg", "notifications": null, "friends_count": 188,
"statuses_count": 325, "profile_background_image_url":
"http://a1.twimg.com/profile_background_images/257176598/hydrangeas_94_
68830.jpg", "default_profile_image": false, "status": { "coordinates":
null, "truncated": false, "created_at": "Tue Jul 05 03:46:03 +0000 2011",
"favorited": false, "id_str": "88091240503058432",
"in_reply_to_user_id_str": "748353", "contributors": null, "text":
"@kmonkeyjam Oh no... I don't know where that bone is but that sounds
terribly painful. How are you still tweeting? Get better!", "id":
88091240503058432, "retweet_count": 0, "in_reply_to_status_id_str":
"87979906226597888", "geo": null, "retweeted": false,
"in_reply_to_user_id": 748353, "source": "Twitter for iPhone",
"in_reply_to_screen_name": "kmonkeyjam", "place": null,
"in_reply_to_status_id": 87979906226597888 }, "display_url": null,
"screen_name": "onesnowclimber", "show_all_inline_media": true,
"following": null } ], "next_cursor_str": "0" }
```

**Figura 2 - Exemplo de uma resposta da API do Twitter.**

Como se pode ver na Figura 2 temos a resposta ao pedido da Figura 1, nesta resposta consegue-se visualizar a lista de “users”, onde se verifica que a lista apenas contém um utilizador, com no nome “Sharon Ly” e toda informação referente a esse perfil.



### 2.2.2. Facebook API

O Facebook não fornece apenas uma API, mas sim um conjunto delas consoante o que se pretende realizar. Estão disponíveis as seguintes API:

- *Live Vídeo API* - permite que os codificadores de vídeo, as câmaras e as aplicações web transmitam *streams* de vídeo diretamente para perfis, páginas e grupos de utilizadores do Facebook.
- *Graph API* - é o núcleo principal das API do Facebook, sendo todas as outras extensões desta. Esta API serve para as aplicações lerem e escreverem no gráfico social do Facebook.
- *Marketing API* - permite criar automatização de tarefas de marketing e realizar a sua gestão.
- *Pages API* - permite criar páginas e fazer a gestão das mesmas.

Antes de fazer o pedido é necessário ter um *token* válido. Para tal pode-se utilizar o código ilustrado no exemplo da Figura 3, com as modificações necessárias:

```
curl -X GET "https://graph.facebook.com/oauth/access_token
?client_id=your-app-id
&client_secret=your-app-secret
&grant_type=client_credentials"
```

Figura 3 - Exemplo de um pedido de um *token* à API do Facebook.

O comando `curl` utilizado nestes exemplos, é uma biblioteca e uma ferramenta de linha de comandos utilizada para transferir dados de ou para um servidor, é nativo dos sistemas Unix, mas também existe para instalação em ambiente Windows [23].

Como se pode ver na Figura 3 o pedido feito pelo `curl` à *Graph API* para receber um *token* válido necessita de 3 parâmetros: “app-id” que representa o identificador único da aplicação já criada, “your-app-secret” representa a password de acesso da aplicação e “client\_credentials” representa a própria palavra e indica a área do serviço.

O exemplo seguinte é da v3.2 da *Graph API*, cujo pedido do id e nome do utilizador é o seguinte, Figura 4:

```
curl -i -X GET \
"https://graph.facebook.com/{your-user-id}
?fields=id,name
&access_token={your-user-access-token}"
```

Figura 4 - Exemplo de um pedido à API do Facebook.

Como se pode ver na Figura 4 o pedido feito à *Graph API*, com o token obtido no exemplo da Figura 3 e irá devolver os campos nome e ID do utilizador.

A Figura 5 representa o tipo de resposta que se obtém com o exemplo da Figura 4, sendo devolvido “Your name” o nome do utilizador e “your-user-id” o id do utilizador.

A resposta é devolvida em formato JSON, como demonstra a Figura 5:

```
{
  "name": "Your Name",
  "id": "your-user-id"
}
```

Figura 5 - Modelo de resposta da API do Facebook do nome e id de um utilizador.

### 2.2.3.Instagram API

De acordo com a própria página do Instagram para *developers*, a API nativa do Instagram foi depreciada, para garantir a privacidade e a segurança dos utilizadores do Instagram [24].De acordo com as datas, estas serão as funcionalidades que estão a ser progressivamente desativadas:

- 31 julho 2018
  - Lista de seguidores – ler a lista de seguidores e utilizadores seguidos
  - Relações – seguir e deixar de seguir contas em nome do utilizador
  - Comentar no perfil publico – comentar e apagar comentários no perfil publico em nome do utilizador
- 11 dezembro 2018
  - Comentar - comentar e apagar comentários em nome do próprio utilizador
  - Conteúdo publico – Ler o perfil publico e media em nome de um utilizador
  - *Likes* – fazer *like/unlike* em nome de um utilizador
  - Subscrição – receber notificações quando é colocado algum tipo de media
- Princípio de 2020
  - Básico – ler o perfil e ficheiros multimédia de um utilizador

Sendo o Instagram agora propriedade do Facebook, foi criada pelo Facebook uma nova API designada Instagram *Graph* API e que é destinada apenas para clientes empresariais/negócios. Para utilização não empresarial deverá ser utilizada então a API antiga, com as limitações já descritas.

#### 2.2.4. Limitações

Em 25 de Maio de 2018 entrou em vigor o Regulamento Geral de Proteção de Dados (RGPD) [25], substituindo a diretiva existente da lei de proteção de dados. Esta nova lei garante aos proprietários dos dados mais controle sobre os mesmos.

Tal facto implicou que as API tivessem de ser alteradas para restringir e limitar algumas pesquisas que até aquela altura poderiam ser feitas. As restrições efetuadas pelas empresas responsáveis tornaram a pesquisa de certos dados públicos mais complicadas ou impossíveis, visto que algumas das informações anteriormente disponibilizadas deixaram de o estar.

Verificou-se recentemente também o crescente número de notícias sobre o acesso indevido aos dados por parte das empresas, com a utilização desses dados para por exemplo influenciar a opinião pública, possivelmente tendo alterado o rumo das eleições nos Estados Unidos da América [26].

Outras notícias dão conta que o Facebook poderá ter influenciado um genocídio em Myanmar [27], levando a que as redes sociais fossem limitando cada vez mais o acesso [28], também influenciadas pela pressão pública exercida sobre elas.

Além dessas limitações, as empresas também foram colocando limites de acessos para pedidos provenientes da API, como se pode verificar pela API do Twitter que limita o número de pedidos a 15 por cada 15 minutos [29]. O mesmo se verifica por exemplo na *Graph* API do Facebook cujo limite é de 200 pedidos por utilizador, por hora [30].

No caso do Instagram o limite é de cerca de 4800 pedidos multiplicado por uma variável que depende do que se pretende aceder, podendo por exemplo ser o número de utilizadores multiplicado pela soma de tempo de uso do processador.

Além do número de pedidos, também se verifica na página de cada API que a quantidade dos dados fornecido é limitado nas versões não pagas, como por exemplo, pedindo os *tweets* de um utilizador à API do Twitter, esta só fornece 200 por cada pedido.

O Instagram deixou de fornecer apoio à sua API nativa e vai consoante o que já foi descrito na secção 2.2.3, desativar algumas funções ao longo do tempo. Como referido anteriormente o Instagram é propriedade do Facebook, tendo o Facebook desativado a API nativa do Instagram após a compra da empresa. No entanto algumas funcionalidades passaram a estar disponíveis na recém criada Instagram Graph API na página do Facebook para *developers* [31], mas igualmente com restrições.

Mais recentemente verificou-se que algumas das ferramentas da API do Facebook foram mesmo desativadas sem aviso prévio, como se pode ler na notícia que indica que a Graph API do Facebook foi alterada para que já não permitisse fazer pesquisas que estariam a ser feitas indevidamente [32].

### 2.3.Noção de Perfil Falso

No âmbito deste projeto considera-se como **perfil falso**, um perfil que, segundo um conjunto de parâmetros e um peso atribuído a cada um deles, não alcança uma certa percentagem de “falsidade”. Cada parâmetro irá incidir sobre informações ou características desse perfil, normalmente um perfil falso contém muito pouca informação ou os dados existentes são valores por omissão.

Também o uso de *botnets* para criação de perfis é algo vulgar e simples, segundo [33], podemos verificar que as botnets conseguem criar perfis falsos mas normalmente utilizam o mínimo necessário para criação de um perfil e quando o fazem em larga escala os perfis criados tem um certo padrão, como por exemplo o número de campos preenchidos.

Pelo descrito anteriormente, consegue-se compreender que tendo selecionado um bom conjunto de parâmetros e atribuindo um peso a cada um deles, refletindo a sua importância, se consiga detetar um perfil falso. Pois como já foi descrito esses parâmetros ou não contém informação associada ou a que contém está por omissão ou é mínima.

Para que o perfil pesquisado seja considerado falso, deverá falhar o máximo de parâmetros possível, por forma a ter uma percentagem de falsidade elevada.

A seguinte fórmula demonstra como é calculada a percentagem de falsidade neste projeto:

$$100 - \left( \sum (Px \times PEx) / n \right) \times 100 = \% \text{ falsidade}$$

$Px$  = representa o valor do parâmetro (*boolean* e que equivalerá a verdadeiro=1 / falso=0), sendo x o número do parâmetro

$PEx$  = representa o peso do parâmetro x

$n$  = representa o número total de parâmetros

Então temos que a percentagem de falsidade é calculada subtraindo ao valor cem o valor obtido, do somatório de todos os valores dos parâmetros multiplicados pelo seu peso, dividindo esse valor pelo número de parâmetros, sendo depois após a divisão multiplicado por cem.

Tomando como exemplo um conjunto de 7 parâmetros, o utilizador do Twitter com o *screen name* “hugobap71416188”, após interrogar a API obtemos a seguinte resposta aos parâmetros:

- P1: Falso, PE1=0.53;
- P2: Falso, PE2=0.85;
- P3: Falso, PE3=0.85;

- P4: Falso, PE4=0.96;
- P5: Falso, PE5=0.917;
- P6: Falso, PE6=1;
- P7: Falso, PE7=0.5;

Resultando o seguinte cálculo, segundo a fórmula descrita anteriormente:

$$100 - (((0*0.53 + 0*0.85 + 0*0.85 + 0*0.96 + 0*0.917 + 0*1 + 0*0.5) / 7) * 100) = 100\%$$

A conta "hugobap71416188" representa um perfil 100% falso, visto que nenhum dos parâmetros considerados tem resposta verdadeira. Ou seja, é considerado falso pois o valor da informação associada aos parâmetros validados é zero ou a que contém não satisfaz os parâmetros necessários.

### 2.3.1. Perfis Falsos nas Redes Sociais

Com o crescimento das redes sociais e do seu número de utilizadores, iniciou-se também um fenómeno de criação de perfis falsos nessas mesmas redes sociais, normalmente com a finalidade de cometer algum tipo de delito e omitir a sua identidade ou tirar proveito financeiro e/ou de popularidade.

Uma investigação feita pelo jornal New York Times em relação a uma empresa designada Devumi evidencia esta mesma realidade. Essa empresa vendia seguidores na rede social Twitter [34]. Segundo o New York Times a empresa detinha cerca de 200 milhões de contas falsas, sendo que 3.5 milhões delas foram criadas automaticamente e poderiam ser vendidas várias vezes.

Na lista de cerca de 200 mil clientes figuravam políticos, atletas e celebridades, que comprariam essas contas para obter mais visibilidade nas redes sociais. Muitas das contas que figuravam na lista de contas detidas pela Devumi continham informações pessoais de verdadeiros perfis das redes sociais, ou seja, algumas das contas seriam clones de contas verdadeiras das redes sociais e continham informação reais tais como fotos, nomes, local de nascimento entre outras informações.

O Twitter não é um caso isolado. Também o Facebook partilha do mesmo problema, sendo que em 2018 lançou um relatório que indica que no primeiro trimestre do mesmo ano desabilitaram cerca de 583 milhões de contas falsas, sem contar com aquelas que não conseguem passar do registo e são desabilitadas logo nesse passo.

O Facebook estima que cerca de 3% a 4% das contas são falsas [35]. Tendo em conta que o Facebook tem 2.23 bilhões de utilizadores ativos mensalmente, tal como referido em cima, temos que 4% desse valor daria um valor de cerca de 892 milhões de contas falsas que ainda estariam ativas no primeiro trimestre de 2018, mesmo após a limpeza que foi realizada pela empresa.

Assim como o Twitter e o Facebook, também o Instagram não foge à regra e também tem perfis falsos na sua rede social. Para ajudar a combater esse problema foram desenvolvidas algumas ferramentas, uma delas “*About This Account*” que permite verificar a informação relativa à conta como data de registo da conta, país onde a conta se localiza, contas com *followers* comuns, *usernames* utilizados no último ano e anúncios que essa conta tem ativos.

Outra ferramenta colocada ao dispor do utilizador, consiste em realizar a validação da conta, para isso o utilizador deve solicitar ao Instagram essa validação, apresentado uma cópia do seu documento de identificação. Na altura da redação deste documento esta ferramenta estava temporariamente desativa.

Uma outra ferramenta consiste na permissão de uso de aplicações de autenticação de dois fatores para login nas contas. Neste caso permitido é o uso do Duo Mobile ou Google Authenticator.

A autenticação de dois fatores fornece uma camada de segurança extra visto que o utilizador necessita de ter conhecimento de dois fatores para poder fazer login, como por exemplo a password para entrada na rede social e por exemplo um código que é enviado para uma aplicação móvel ou email.

### **2.3.2. Crime com recurso às redes sociais**

Cibercrime, tal como a própria palavra indica é o crime cometido através da comunicação entre redes de computadores, como por exemplo Internet [36].

Com o crescente aumento da utilização das redes sociais, também o mundo do crime começou a voltar-se para as redes sociais e os criminosos começaram a arranjar os mais diversos esquemas para conseguirem vingar no mundo virtual e tirar disso proveito financeiro. Podemos ter como referência o jornal Guardian que indica que no Reino Unido de 2008 para 2012 houve um aumento do crime nas redes sociais Facebook e Twitter de 780% [37].

Segundo Annabelle Graham nos últimos 6 anos houve um aumento de cerca de 3 mil milhões de registos de ocorrências de cibercrime [38], sendo que não tem tendência para abrandar. De acordo com um estudo da Universidade de Birmingham existem seis tipos de criminosos que utilizam o Facebook para cometer crimes, designadamente: reator, informador, antagonista, fantasista, predador e impostor [39].

O reator é o tipo de criminoso que reage de forma violenta a algo que o enfurece, o informador é aquele que informa que cometeu ou quer cometer um certo crime, o antagonista fala e responde de forma agressiva no Facebook transpondo depois disso para a realidade. O predador cria um perfil falso para atrair a vítima e encontrar-se pessoalmente com ela, os fantasistas são aqueles que tem a dificuldade em aceitar a realidade e que cometem o crime para tentar manter a realidade de que tudo está bem e por último o impostor que é o criminoso que se faz passar por outra pessoa.

Tal como no resto do mundo, também Portugal não foge à regra e o cibercrime é uma preocupação crescente. Podem ler-se vários apelos das forças de segurança alertando para o cibercrime, mais concretamente para o que utiliza como meio de disseminação ou de prática as redes sociais [40][41].

Segundo o relatório anual do gabinete de cibercrime que presta informações sobre o período de janeiro de 2013 até junho de 2016, poderá ler-se que as autoridades portuguesas enviaram cerca de 2.622 pedidos de esclarecimentos ao Facebook, 2847 à Microsoft e 2798 à Google [21][22]. No relatório que compreende as datas de 2015 a 2017 é evidenciado o crescente aumento do cibercrime com especial referência no ponto 8 B “Cibercriminalidade”, em que refere “um grande significado estatístico, as participações respeitantes à criação de perfis falsos em redes sociais (em particular no Facebook)”[43].

Pode ler-se também nos media , várias notícias sobre crimes praticados com recurso às redes sociais, podendo dar como exemplo a criação de perfis falsos para difamação que segundo Mariana Oliveira do Jornal o Público teria dado em 2014 origem a cerca de 1000 denúncias às autoridades[1]. Também a usurpação de uma conta de Facebook já chegou à barra do tribunal [44], assim como chantagem e divulgação de fotos e vídeos íntimos [45]. Já em 2015 e segundo o jornal Público a Polícia Judiciária tinha registado cerca de 50 denúncias de sextorsion, que é uma forma de conseguir extorquir favores sexuais às vítimas usando a coação de forma não física [46].

Mostrando que este fenómeno não se delimita às grandes cidades podemos ler no Jornal de Notícias onde a inspetora chefe, Fátima Marques da Polícia Judiciária de Leiria indica que todos os dias recebe entre 3 a 4 queixas de crime informático evidenciando que alguns desses crimes são feitos com recurso às redes sociais e perfis falsos[47], fala também de *ciberbullying* ou seja uma prática de agressão moral praticada na internet por um grupo em relação a uma pessoa.

Recentemente houve vários escândalos relacionados com a exfiltração de dados, que abalaram a opinião pública. Um dos mais mediáticos é o caso Cambridge Analytica [48], em que segundo as notícias cerca de 50 milhões de utilizadores do Facebook viram os seus dados roubados pela empresa Cambridge Analytica. Após a sua recolha os dados eram analisados e eram criados anúncios personalizados de forma a que aqueles que fossem eleitores nos EUA fossem guiados para esses anúncios, influenciando o seu voto para votar em Donald Trump, que viria a ganhar posteriormente as eleições.

Em Portugal também uma notícia apresentada pelo Diário de Notícias apresenta informação sobre uma alegada campanha difamatória feita por 18 perfis falsos na rede social Twitter, contra o atual governo e alguns candidatos do PS [49]. Essa notícia deu origem a uma queixa à Comissão Nacional de Eleições (CNE) e à Entidade Fiscalizadora do Financiamento dos Partidos (EFFP) [50].

## **2.4.Estado da Arte - Detecção de Perfis Falsos nas Redes Sociais**

Sendo um problema crescente e com grande impacto na sociedade atual, existem já algumas tentativas para a detecção automática de perfis falsos nas redes sociais. Tal como foi descrito anteriormente, até mesmo as empresas detentoras dessas redes sociais implementaram ferramentas para tentar minimizar o número de perfis falsos. Existem várias aproximações propostas à resolução do problema por vários investigadores.

Em 2012 surgiu uma abordagem que consiste em fazer um estudo das redes sociais e caracterizar vários perfis de utilizadores. Para tal foram recolhidas várias informações para depois identificar perfis padrão. Em seguida estudaram-se os utilizadores com base nesses perfis e verificaram-se os desvios em relação ao padrão esperado. Os principais parâmetros por eles evidenciados foram o número de amigos ao longo do tempo, a interação na rede social e a evolução gráfica da estrutura da rede social ao longo tempo [51].

Em 2013, tendo em conta o problema da clonagem de perfis, surgiu uma pesquisa que teve como base a criação de uma ferramenta que permitiria às empresas gestoras das redes sociais instalar nos seus servidores uma ferramenta que, com base no endereço IP dos utilizadores e num algoritmo, avaliaria se uma conta poderia ser um clone [52]. Para isso procediam da seguinte forma: de cada vez que um utilizador iniciava uma sessão na rede social, eram guardados os 16 bits mais significativos do endereço IP e colocado num buffer com tamanho 4.

De seguida, de cada vez que um pedido de amizade era feito, era verificado se na lista de amigos da pessoa a que o pedido era dirigido, já existia alguma pessoa com esse nome e se não existisse o procedimento parava. Por cada pessoa existente na lista de amigos com o mesmo nome era então feito um cálculo de similaridade com base num algoritmo já existente e em alguns atributos, como género, localidade, etc.

Caso o valor dessa similaridade fosse superior a um valor pré-definido, indicava que a conta poderia ser um clone ou ser uma outra conta criada pela mesma pessoa. Para verificar se a conta era um clone potencialmente criada por um atacante, comparava-se o IP de ambas as contas e caso o prefixo da sequência IP fosse diferente indicava que potencialmente a conta seria clonada.

Em 2015 uma investigação focada nas contas falsas de utilizadores do Twitter, utiliza uma abordagem baseada num padrão de detecção de características com base na atividade dessa conta. Para tal foi usada a API disponibilizada pelo Twitter e foi recolhida informação sobre 33 características de 62 milhões de contas cuja informação estava disponível ao público [53].

Seguidamente, como verificaram que muitas dessas características não tinham sido preenchidas pelos utilizadores, fizeram uma pesquisa semianual e selecionaram os seguintes



atributos: id, número de seguidores, número de amigos, verificação de conta, data criação, descrição, localização, data de atualização, URL da imagem de perfil e nome de perfil.

Com base na lista anterior criaram grupos de contas que tivessem igual: nome; descrição e localização. Nesses grupos foi depois aplicado um algoritmo de reconhecimento de padrões ao *screen name*. O algoritmo de reconhecimento de padrões é um algoritmo que classifica algo quando comparado com um padrão, neste caso a similaridade do *screen name* com base em outro *screen\_name*.

Dentro de cada grupo foi utilizado o primeiro *screen name* e comparado com os restantes, se o valor fornecido pelo algoritmo fosse superior ao definido pelo estudo (0,1) era considerado similar e colocado num grupo, os restantes foram colocados em outro grupo e levados novamente a análise, e assim sucessivamente até serem todos analisados.

Os grupos classificados como contas com *screen name* similares foram depois identificados com o uso de expressões regulares. As expressões regulares identificam dentro de um texto se existe um determinado padrão, neste caso se um *screen name* tivesse pelo menos 4 caracteres seguidos iguais a outro *screen name*, se isso acontecesse eram então classificadas essas contas como falsas.

Mais recentemente, em 2018, uma equipa de investigadores das universidades de Israel e Washington desenvolveram um algoritmo genérico que permite detetar contas falsas na maioria das redes sociais incluindo Facebook e Twitter [54]. De acordo com o estudo efetuado, este método de deteção tem como pressuposto que as contas falsas tendem a estabelecer ligações improváveis com outros utilizadores da rede.

Para detetar essas ligações recorrem a um novo algoritmo genérico e não supervisionado e topologia gráfica. O algoritmo consiste em 2 iterações principais. A primeira consistiu em criarem um classificador para prever uma ligação baseado apenas na topologia gráfica. A segunda iteração, consistiu em gerar uma série de características com base no prognóstico criado anteriormente. Seguidamente utilizando as características geradas construíram um classificador de contas falsas.

Todas as investigações apresentadas anteriormente representam diferentes formas de tentar dar resposta a um problema comum, mas nenhuma delas é comparável com o que se desenvolvemos neste projeto. A investigação que mais se aproxima e que servirá de base para o trabalho realizado foi elaborada por Ahmed El Azab, Amira M. Idrees, Mahmoud A. Mahmoud, Hesham Hefny em 2015 [55]. A metodologia aí apresentada foi desenhada para o Twitter mas, segundo os autores, pode servir de base para aplicação noutras redes sociais. Segundo [55] a análise às contas do Twitter é feita com base num conjunto de parâmetros e uma ponderação atribuída a cada um.

A abordagem apresentada em [55] é bastante diferente das apresentadas anteriormente, as outras basearam-se em gráficos ou algoritmos para resolver o problema, como por exemplo a abordagem que tem como base o IP [18], que facilmente pode ser ludibriada alterando o IO ou usando uma VPN.

Em relação às abordagens baseadas em gráficos, que utilizam as ligações entre utilizadores como base para o estudo da falsidade das contas [56][57][58], podemos ver que também é possível criar facilmente perfis falsos. Para isso poder-se-ia utilizar *botnets* fazendo-se passar por utilizadores verdadeiros, passando assim pelas validações base das redes sociais. Outra das formas seria usando engenharia social fazendo-se amigo de muitos utilizadores, inviabilizando assim a utilização de uma deteção baseada em análise gráfica.

A abordagem de Ahmed El Azab, Amira M. Idrees, Mahmoud A. Mahmoud, Hesham Hefny tem como principal objetivo analisar as contas e atribuir um resultado em termos de falsidade, baseando-se no mínimo conjunto de parâmetros possível. Para isso, a abordagem divide-se em dois pontos principais: o primeiro foi encontrar os fatores que evidenciem que um perfil é falso; o segundo ponto foi aplicar um algoritmo de classificação que utilizando os fatores encontrados anteriormente descubra as contas falsas.

Segundo os investigadores, em anteriores pesquisas [59][55][60], foram utilizados um elevando número de parâmetros desnecessariamente, visto que muitos deles não eram utilizados pelos utilizadores das redes sociais, ou possuíam valores por omissão. O conjunto de parâmetros utilizados por outros investigadores está representado na Figura 6.

SET OF ATTRIBUTES PROPOSED BY DIFFERENT RESEARCHERS			
[8]	[9]	[10]	[15]
Number Of Followers Number Of Followees Fraction Of Followers Per Followees Number Of Tweets Age Of The User Account Number Of Times The User Was Mentioned Number Of Times The User Was Replied To Number Of Times The User Replied Someone Number Of Followees Of The User's Followers Number Tweets Received From Followees Existence Of Spam Words On The User's Screen Name The Minimum Of The Time Between Tweets The Maximum Of The Time Between Tweets The Average Of The Time Between Tweets The Median Of The Time Between Tweets Number Of Tweets Posted Per Day Number Of Tweets Posted Per Week	<ul style="list-style-type: none"> <li>Followers_Count</li> <li>Id, Friends_Count</li> <li>Verified</li> <li>Created_At</li> <li>Description</li> <li>Location</li> <li>Updated</li> <li>Profile_Image_Url</li> <li>Screen_Name</li> </ul>	<ul style="list-style-type: none"> <li>FF ratio (R): following / followers ((where following, in the Twitter jargon, is the number of friend requests sent, and followers is the number of users who accepted the request) Large for spammers</li> <li>URL ration = U = messages containing urls / total messages.</li> <li>Message Similarity:</li> <li>Similarity among the messages sent by a user.</li> <li>Friend Choice (F) = <math>F = T_n / D_n &gt; 1</math> for spammers, where <math>T_n</math> is the total number of names among the profiles' friend, and <math>D_n</math> is the number of distinct first names.</li> <li>Messages Sent (M): We use the number of messages sent by a profile as a feature</li> <li>Spammers <math>M &lt; 20</math> message</li> <li>Friend Number:</li> <li>number of friends a profile has = thousands for humans and few for spammers</li> </ul>	<ul style="list-style-type: none"> <li>The Ratio Friends Followers Of The Account Under Investigation Is 50:1, Or More;</li> <li>More Than 30% Of All The Tweets Of The Account Use Spam Phrases, Such As "Diet", "Make Money" And "Work From Home";</li> <li>The Same Tweets Are Repeated More Than Three Times, Even When Posted To Different Accounts;</li> <li>More Than 90% Of The Account Tweets Are Retweets;</li> <li>More Than 90% Of The Account Tweets Are Links;</li> <li>The Account Has Never Tweeted;</li> <li>The Account Is More Than Two Months Old And Still Has A Default Profile Image;</li> <li>The User Did Not Fill In Neither Bio Nor Location And, At The Same Time, She Is Following More Than 100 Accounts.</li> </ul>

Figura 6 - Conjunto de parâmetros proposto por diferentes investigadores [54].

Nesta investigação começaram com 22 parâmetros, ilustrados na Figura 7, que foram testados para verificar quais desses 22 permitiriam ter uma maior precisão na identificação de uma conta falsa.

PROPOSED ATTRIBUTES AND DETERMINED WEIGHT	
Attributes	Weight
the account has at least 30 followers	0.53
the account has been geo-localized	0.85
it has been included in another user's favorites	0.85
it has used a hashtag in at least one tweet	0.96
it has logged into Twitter using an iPhone	0.917
a mention by twitter user	1
it has written at least 50 tweets	0.01
it has been included in another user's list	0.45
(2*number followers) _ (number of friends)	0.5
User have at least one Favorite list	0.17
the profile contains a name	0.0
the profile contains an image	0.0
the profile contains a biography	0.0
the profile contains a URL	0.0
it writes tweets that have punctuation	0.0
it has logged into Twitter using an iPhone	0.0
it has logged into Twitter using an Android device	0.0
the profile contains a physical address	0.0
it has logged into twitter.com website	0.0
it is connected with Foursquare;	NA
it is connected with Instagram	NA
it has logged into Twitter through different clients	NA

Figura 7- Conjunto de 22 parâmetros utilizado na investigação.

Para validar qual o número mínimo de parâmetros e quais os parâmetros a utilizar os investigadores seguiram o seguinte método de trabalho:

- Primeiro arranjaram um *dataset* de contas do Twitter disponível no projeto “the Fake project” [61], em que o *dataset* tem 1481 contas reais de um antigo projeto designado #elezioni2013. De seguida recolheram-se mais 469 de utilizadores reais e a estas contas somaram-se mais 3000 contas falsas disponíveis na Internet.
- Após isso, criou-se um conjunto de 22 parâmetros (Figura 7). Com esses parâmetros recolheram-se informações sobre as contas do *dataset* especificados anteriormente.
- De seguida utilizaram os 5 algoritmos de aprendizagem para classificação : Random Forest [62], Decision Tree [63], Naïve Bayes [64], Neural Network [65] e Support Vector Machine [66] para *k-fold cross-validation*. O método *k-fold* consiste em dividir o *dataset* em k partes, usando k-1 partes para treino e a parte remanescente para teste, fazendo isso k vezes. Em cada uma das k vezes, testa-se o modelo com um *fold* diferente calculando a métrica escolhida para avaliação do modelo.
- Utilizando os 22 parâmetros utilizaram o método explicado anteriormente testaram o modelo e compararam resultados.

- Dos 22 parâmetros escolheram-se 19, repetiram-se os mesmos testes e a análise efetuada aos 22 parâmetros.
- Após isso foi feita uma comparação de resultados com os obtidos anteriormente. Usando o algoritmo de Gain Measure [67] calcularam uma ponderação para os 22 parâmetros (Figura 7), por forma a tentar obter o melhor resultado com o mínimo de parâmetros necessários.
- Voltaram a repetir os testes descritos anteriormente usando o método *k-fold*, e foi feita a comparação de resultados. Procederam de igual forma para o conjunto dos 19 parâmetros.

De acordo com [68] foram escolhidos 7 parâmetros dos 22 iniciais, sendo os 7 parâmetros aqueles cuja ponderação era igual ou superior a 50%. Com esses 7 parâmetros foi repetido o conjunto de testes usando o método *k-fold* e foi feita uma comparação de resultados tal como anteriormente. Para o conjunto dos 19 parâmetros foi repetido o mesmo procedimento descrito anteriormente, escolhendo-se aqueles que tinham uma ponderação igual ou superior a 50% e feito os testes usando o método *k-fold*. Neste caso apenas resultou a escolha de 6 parâmetros.

Para avaliar cada algoritmo utilizaram-se os seguintes quatro indicadores: TN = *True Negative*, TP = *True Positive*, FN = *False Negative*, FP = *False Positive*. Foram calculadas as seguintes medidas de avaliação: *Precision* (Figura 8), *Recall* (Figura 9) .

$$Precision = \frac{Verdadeiros\ Positivos\ (TP)}{Verdadeiros\ Positivos\ (TP) + Falsos\ Positivos\ (FP)}$$

Figura 8 - Fórmula da Precision [70].

Tal como pode ver-se representado na Figura 8 a fórmula da *Precision* é o resultado da divisão do dividendo Verdadeiros Positivos (TP) pelo divisor, que é a soma dos Verdadeiros Positivos (TP) com Falsos Positivos (FP). A precisão serve para indicar daqueles que classificados como verdadeiros quantos são efetivamente verdadeiros.

Neste caso o dividendo seria o número de contas detetadas como falsas e que efetivamente o são, o divisor é número de contas detetadas como falsas e que efetivamente o são, com a soma das contas que foram classificadas como falsas, mas na realidade são verdadeiras.

$$Recall = \frac{Verdadeiros\ Positivos\ (TP)}{Verdadeiros\ Positivos\ (TP) + Falsos\ Negativos\ (FN)}$$

Figura 9 - Fórmula da Recall [70] .

Na Figura 9 encontra-se representada a fórmula da *Recall*, que exprime a frequência com que o classificador encontra os exemplos de uma classe, neste caso, seria com que frequência uma conta falsa é classificada como sendo falsa.

Aplicando a fórmula, seria a divisão do dividendo Verdadeiros Positivos (TP) neste caso contas falsas que efetivamente o são, pelo divisor Verdadeiros Positivos (TP) contas falsas que efetivamente o são, somando com os Falsos Negativos (FN) contas indicadas como verdadeiras, quando efetivamente não o são.

Após utilizar o *dataset* do projeto “*The Fakeproject*” e terem utilizado o algoritmo do *Gain measure* a cada parâmetro e de acordo com a sua importância sobre o *dataset*, concluíram que dos 22 parâmetros apenas 10 possuíam uma ponderação superior a 0%. Os parâmetros e a sua ponderação podem ser consultados na Figura 10.

CLASSIFYING ATTRIBUTES WITH THEIR CORRESPONDING WEIGHTS	
Attributes	Weight
the account has at least 30 followers	0.53
the account has been geo-localized	0.85
it has been included in another user's favorites	0.85
it has used a hashtag in at least one tweet	0.96
it has logged into Twitter using an iPhone	0.917
a mention by twitter user	1
it has written at least 50 tweets	0.01
it has been included in another user's list	0.45
(2*number followers) _ (number of friends)	0.5
User have at least one Favorite list	0.17

Figura 10 - Classificação dos 10 parâmetros e seu peso.

Seguidamente na Figura 11 estão os parâmetros que utilizando o mesmo *dataset* mencionado anteriormente, tiveram uma ponderação igual ou superior a 50%.

CLASSIFYING ATTRIBUTES HAVING WEIGHT ABOVE OR EQUAL 50% WITH THEIR CORRESPONDING WEIGHTS	
Attributes	Weight
the account has at least 30 followers	0.53
the account has been geo-localized	0.85
it has been included in another user's favorites	0.85
it has used a hashtag in at least one tweet	0.96
it has logged into Twitter using an iPhone	0.917
a mention by twitter user	1
(2*number followers) _ (number of friends)	0.5

Figura 11 - Classificação dos 7 parâmetros com ponderação igual ou superior a 50%.

Com base na experiência descrita e nos resultados obtidos, os investigadores conseguiram encontrar os parâmetros mínimos necessários para identificar contas falsas (Figura 11), tendo em conta a metodologia definida.

Em relação ao projeto e metodologia que pretendo desenvolver, as redes sociais que este pretende abranger e em comparação com o estado da arte já descrito anteriormente, posso concluir que não existe nada similar. Como já foi descrito a abordagem que mais se aproxima é a [55], a qual, apresenta um conjunto de parâmetros mínimos estudados e validados para a identificação de perfis falsos no Twitter, por esse motivo servirá de base para minha solução.

Não existe no mercado nenhuma plataforma que, com base num conjunto de parâmetros, devolva uma percentagem que indique que um determinado perfil pesquisado seja potencialmente falso.

Existem alguns estudos efetuados tal como já foi descrito com outro tipo de abordagens para tentar determinar a veracidade de um perfil em algumas das redes propostas, mas nenhum contemplou a elaboração de algum tipo de aplicação para elaborar essa análise seguindo a metodologia que se utilizou no âmbito deste trabalho.

Em termos de aplicações, as que mais se aproximam são as da IntelTechniques [71] desenvolvidas por David Westcott e Michael Bazzell. São ferramentas que permitem questionar a API das redes sociais com base no ID da conta, não dando informações sobre a falsidade do perfil, mas recolhendo informação pública sobre esse perfil. A IntelTechniques dedica-se às seguintes atividades: *OSINT training*, *Privacy consulting* e *digital security*.

Como a própria sigla indica, OSINT (*Open Source INTelligence* ou Inteligência de Fontes Abertas), modelo de inteligência que visa encontrar, selecionar e adquirir informações de fontes públicas e analisá-las para que junto com outras fontes possam produzir um conhecimento. Neste caso as informações recolhidas são provenientes das redes sociais e estão disponíveis para que qualquer pessoa possa pesquisar, é a chamada informação pública.

As ferramentas que a IntelTechniques criou e que mais se aproximam à temática em estudo são as seguintes: “Custom Instagram Search Tools” [72], “Custom Twitter Tools” [73] e “Custom Facebook Tools” [74]. Todas as redes sociais escolhidas apresentam API que podem ser questionadas e devolvem os resultados dessa pesquisa, principalmente no formato JSON.

As ferramentas da IntelTechniques mencionadas anteriormente, funcionam da seguinte forma: primeiramente fazem o pedido à API específica de cada rede com o ID da conta ou perfil; a API dessa rede social devolve a informação em JSON; e depois a ferramenta irá apresentá-la ao utilizador de uma forma mais “amigável”, ou seja, não serializada.

Desde 6 junho de 2019 as ferramentas da IntelTechniques já não estão disponíveis ao público. Após o website ter recebido um conjunto de ataques em larga escala, foi posteriormente suspenso, sob a ameaça de imposições legais ao autor. Tal como explica num *podcast* [75], o autor resolveu retirar as ferramentas assim como outros recursos disponíveis ao público para poder ter novamente o website online e não ter problemas judiciais.

## 3. Arquitetura

No presente capítulo descreve-se a abordagem utilizada, os parâmetros estudados para a rede social Twitter, o desenho de alto nível do funcionamento da solução e o *workflow*, assim como algumas considerações a ter em conta no seu funcionamento.

Inicialmente, o presente estudo previa que a aplicação abordasse as redes sociais Twitter, Facebook e Instagram e que fosse possível tendo em conta a metodologia aplicada, identificar perfis falsos em cada uma delas.

Infelizmente devido a limitações recentemente impostas pelo Facebook e Instagram nas suas API, já descritas no capítulo 2.2.4 não é possível recolher a informação referente às contas, a não ser da própria conta do utilizador. Pelos motivos apresentados o desenvolvimento apenas visou a rede social Twitter.

### 3.1. Abordagem

Inicialmente foi necessário fazer uma extensiva pesquisa sobre o que já existia e que poderia ser o ponto de partida para definir a abordagem. Após essa pesquisa já descrita no capítulo 2.4, verificou-se que a melhor opção para a rede social Twitter seria a dos autores Ahmed El Azab, Amira M. Idrees, Mahmoud A. Mahmoud, Hesham Hefny [55].

Tendo como base o estudo já referido e a rede social a utilizar, foi necessário estudar e escolher os parâmetros a utilizar. Após a escolha dos parâmetros, seguiu-se a definição da arquitetura e a metodologia a utilizar no desenvolvimento da aplicação.

A metodologia poderá ser aplicada a qualquer rede social, tendo em conta o uso de parâmetros específicos para cada rede e respetivos pesos.

Como foi necessário recolher informações da rede social Twitter foi feito um estudo prévio de funcionamento da API. Para poder utilizar a API do Twitter é obrigatório criar uma conta na página de *developers* do Twitter, através do endereço <https://developer.twitter.com> e solicitar acesso às API. Caso não tenha uma conta no Twitter é necessário criá-la primeiro.

Após ter conta de *developer* é necessário criar uma aplicação para ter acesso às *keys* e *tokens* que permitem fazer chamadas à API e obter as correspondentes respostas. Uma *key* é uma credencial única de acesso que identifica a aplicação criada. Um *token* é uma chave digital para acesso com essa aplicação.

Tendo em conta a metodologia e o conhecimento que possuo em desenvolvimento de aplicações, fiz a escolha da linguagem de programação tendo escolhido a que tenho mais conhecimentos e desenvolvi a aplicação. Tendo em conta os parâmetros, são feitos pedidos via API à rede social. Caso esses pedidos devolvam uma resposta, esta é analisada e classificada com um valor. No final é aplicado o peso a cada um dos parâmetros tendo em conta o valor anterior e é calculado o valor da probabilidade de falsidade do perfil.

Após elaborar a aplicação, criei um *dataset* de testes, elaborei um cenário de teste e recolha de resultados. No final disponibilizei a aplicação ao público no endereço: [www.fakeprofiles.online](http://www.fakeprofiles.online).

De acordo com Ahmed El Azab, Amira M. Idrees, Mahmoud A. Mahmoud, Hesham Hefny [55], o conjunto de parâmetros que apresenta melhores resultados são os já apresentados na Figura 10 e Figura 11, tendo sido esses que utilizei para realizar os primeiros testes, assim como os pesos.

Além dos 10 parâmetros e 7 parâmetros apresentados pelo estudo já referido, resolvi acrescentar um novo parâmetro que na altura do estudo de Ahmed El Azab ainda não existiam dados que o pudessem fundamentar. Trata-se do parâmetro “Validação da conta” e consiste na validação da conta efetuada pelo Twitter, após o utilizador solicitar a validação, submeter um documento de identificação e preencher um conjunto mínimo de dados.


Sendo um parâmetro possivelmente importante atribui-lhe o peso de 1, contudo no capítulo 4.4 iremos verificar o impacto desse parâmetro no resultado e verificar se ele é mesmo um parâmetro importante e influenciador do resultado.

Para facilitar o entendimento dos parâmetros, descrevo de seguida alguns termos e a sua designação para a rede social Twitter:

- **Membro/utilizador:** quando uma pessoa se regista na rede social, passando a ter uma conta;
- **Perfil:** após registo o membro passa a ter um perfil e uma página, a sua informação como fotografia, nome, localidade, data de início, etc. passa a estar disponível aos restantes membros;
- **Seguidor:** um utilizador pode optar por seguir outro unilateralmente, sendo assim seu seguidor (*follower*), têm acesso ao perfil e tweets da outra pessoa;
- **Conta Favorita:** é um perfil que pelo menos é seguido por outro perfil, tornando-se assim uma conta favorita desse último;
- **Hastag:** é uma palavra precedida por o símbolo “#”;
- **Menção:** uma menção é feita com recurso a um *tweet* e neste se identifica outro perfil no comentário, para mencionar um perfil utiliza-se o símbolo “@” antes de colocar o *screen name* desse perfil;
- **Lista:** Uma lista de favoritos é constituída por um grupo de contas do Twitter. Um utilizador pode criar suas próprias listas ou inscrever-se em listas criadas por outros, ficando assim seguidor dessas contas.

Descrevem-se de seguida os 11 parâmetros utilizados e uma breve descrição de cada um deles:



- **A conta deverá ter no mínimo 30 seguidores:** o perfil pesquisado deve ter no mínimo 30 seguidores;
- **A conta deve estar geolocalizada:** o perfil deve ter definido a sua localização espacial, desta forma conseguiremos saber onde se encontra localizado fisicamente essa pessoa;
- **A conta deverá ser favorita de outro utilizador:** o perfil pesquisado deverá ser seguido pelo menos por outra pessoa;
- **Pelo menos uma hashtag num tweet:** o perfil pesquisado deverá ter pelo menos um hashtag num tweet;
- **Pelo menos um login com um iphone:** o perfil pesquisado deverá ter feito um login pelo menos uma vez com um iphone;
- **Uma menção feita por um utilizador do Twitter:** o perfil pesquisado ter sido mencionado por outro perfil. Como por exemplo um tweet escrito por mim mencionando Donald Trump: “O perfil do Presidente dos Estados Unidos é @realDonaldTrump”;
- **Pelo menos 50 tweets:** o perfil ter escrito pelo menos 50 tweets;
- **Estar incluído numa lista de outro utilizador:** o perfil pesquisado deverá fazer parte de uma lista criada por outro perfil;
- **(2\* número de followers)> (número de perfis seguidos):** neste caso o perfil pesquisado deverá ter no mínimo duas vezes mais seguidores, que o número de perfis que ele segue;
- **Ter pelo menos uma lista de favoritos:** neste parâmetro o perfil pesquisado deverá ter criado no mínimo uma lista de favoritos;
- **Conta estar validada:** o emblema azul de verificação no Twitter  permite que as pessoas saibam que a conta é válida ou seja os dados são reais. Para ter a conta validada pelo Twitter é necessário ter alguma informação introduzida na conta e requerer a sua validação mediante o envio de documento de identificação[76]. Após análise o Twitter indica se a conta é autêntica e atribui-lhe o referido emblema como distinção e prova.

### 3.2.Desenho da Arquitetura

A solução desenvolvida é constituída por uma aplicação que irá questionar a rede social Twitter por via da API. Essa API devolve uma resposta em JSON. A resposta é interpretada e analisada na plataforma e caso seja válida, é-lhe atribuído um valor. Esse valor juntamente com o peso permitirá calcular o valor final que depois é utilizado no cálculo da percentagem de falsidade da conta.

Com base na solução técnica desenvolvida, decidi colocar num alojamento web a aplicação com o nome “fakeprofiles”. O endereço [www.fakeprofiles.online](http://www.fakeprofiles.online) é um endereço de fácil perceção, memorização e compreensão.

A Figura 12 ilustra o esquema de alto nível de funcionamento da solução, onde é possível verificar o fluxo de informação entre a plataforma e os demais intervenientes. Apresento a descrição dos pontos da Figura 12 :

1. O Utilizador acede ao endereço [www.fakeprofiles.online](http://www.fakeprofiles.online) insere o *screen name* do perfil a pesquisar e submete o pedido.
2. A aplicação recebe o pedido do utilizador e faz um pedido de autorização/validação de *Key* e *Token* ao servidor responsável pela Autorização do Twitter.
3. O servidor de Autorização questiona a BD para verificar se a *Key* e *Token* existem e são válidos.
4. Resposta da BD ao pedido de descrito no ponto 3.
5. Resposta do servidor de Autorização à aplicação.
6. A aplicação submete pedidos à API para dar resposta aos parâmetros.
7. API quando recebe os pedidos feitos pela aplicação, para lhe dar resposta faz pesquisa da informação solicitada à BD.
8. Resposta com a informação solicitada no ponto 7.
9. Resposta por parte da API aos pedidos à API no ponto 6.
10. Após análise das respostas recebidas, e do cálculo da % de falsidade a aplicação guarda na BD o resultado para servir como log.
11. Resposta ao utilizador via browser com a % de falsidade ou caso algo ocorra fora do esperado com mensagem de erro.

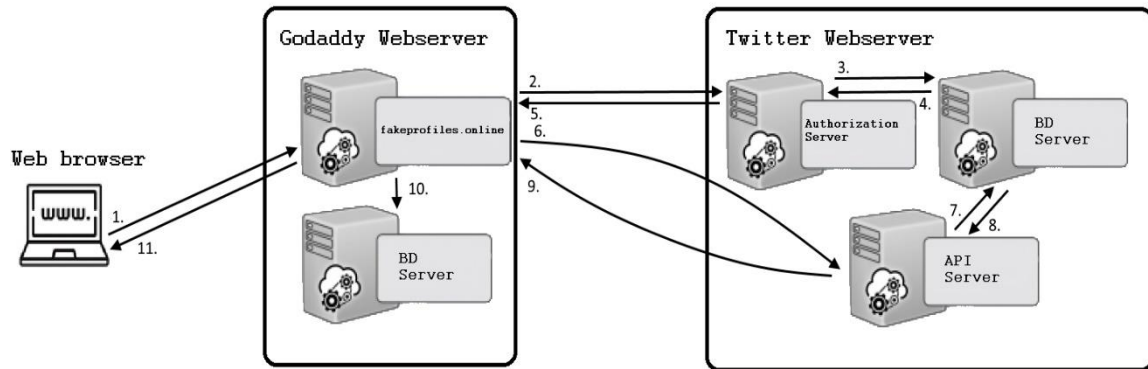


Figura 12 - Esquema de Alto Nível de funcionamento da solução.

A Figura 13 ilustra o *workflow* mais detalhado da solução e do seu funcionamento.

Sendo o funcionamento o seguinte: o utilizador insere na aplicação o “*screen name*” do perfil. De seguida a plataforma irá verificar se existe uma conta com esse “*screen name*”. Se existir, continua o processamento. Caso contrário termina a sua pesquisa e informa o utilizador.

No caso de existir um utilizador com o “*screen name*” introduzido, para cada parâmetro a aplicação irá questionar a API da rede social em causa. Quando esta receber a resposta, a aplicação irá verificar se a resposta é válida. Se for válida é atribuído um valor à resposta e segue para o próximo parâmetro; caso contrário a pesquisa termina e informará o utilizador.

Após ter percorrido todos os parâmetros e ter todas as respostas, é calculado o valor de cada parâmetro consoante o peso. No final é calculada a percentagem de falsidade da conta com base no valor anteriormente calculado de cada parâmetro. Guarda os valores dos parâmetros e percentagem de falsidade, como log na BD. Para terminar é apresentada a percentagem de falsidade ao utilizador.

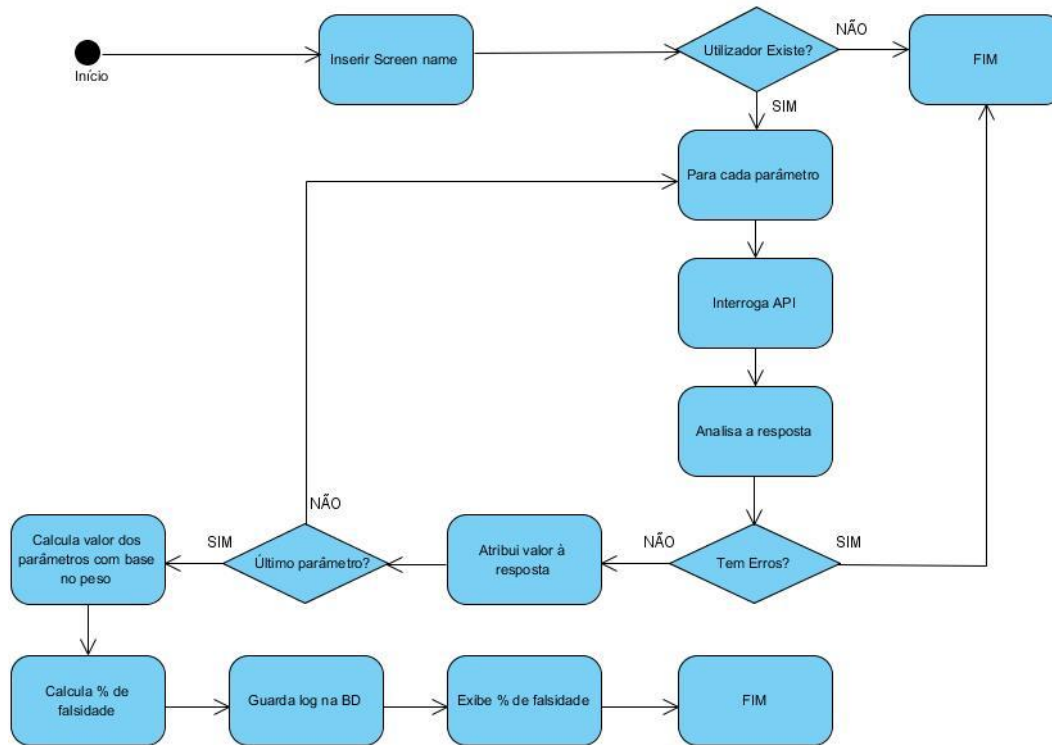


Figura 13 - Esquema de funcionamento detalhado da solução.

### 3.3. Considerações sobre o funcionamento

Ao longo da elaboração deste projeto, tanto na parte teórica como prática foram aparecendo algumas condicionantes que tiveram algum impacto na aplicação, que seguidamente irei descrever.

Durante o estudo teórico verifiquei que como não iria utilizar a versão das API pagas das redes sociais, o acesso aos dados iria ser inferior. Mais concretamente o número de acessos e a frequência com que poderia fazer pedido a essas API, também as API pagas tem mais funcionalidades que as grátis.

Um exemplo do descrito anteriormente e mencionado no capítulo 2.2.4 pode ser verificada na aplicação desenvolvida, pesquisando na rede Twitter mais de 15 contas em menos de 15 minutos, o resultado será o barramento de acesso durante 10 minutos, só após esse tempo será de novo possível o acesso.

Em termos de acesso aos dados as limitações também existem, por exemplo no número de *tweets* que se conseguem recolher. Só é possível recolher 200 *tweets* por cada pedido à API e num máximo de 3600 no total através do recurso a vários pedidos.

Foram feitos testes para o Facebook e Instagram, mas devido à aplicação sucessiva de restrições às API, já descritas no capítulo 2.2.4 para cumprir o RGPD, tal como para colmatar sucessivas falhas de privacidade exploradas por várias empresas, tornou-se impossível segundo os testes efetuados aceder à informação necessária. Pelo descrito anteriormente, a aplicação apenas dispõe da rede social Twitter para fazer pesquisas.

Outra limitação encontrada no Twitter tem a ver com os perfis de privacidade que o utilizador pode seleccionar. Caso o utilizador limite o acesso aos dados, deixando assim de serem informação pública, não é possível aferir sobre a falsidade do perfil. Essa opção encontra-se nos "Settings -> Privacy and Safety -> Privacy -> Tweet Privacy", opção "Protect your Tweets". Caso isso aconteça, a aplicação não fará a pesquisa pois não consegue recolher dados.

### **3.4.Desenvolvimento**

Tal como já foi descrito, optei por desenvolver uma aplicação web. A aplicação foi desenvolvida em C# utilizando a framework .NET. A aplicação está localizada no seguinte endereço web: [www.fakeprofiles.online](http://www.fakeprofiles.online)

O utilizador do website ao aceder e navegar no site está a acordar com a política de privacidade e cookies, de acordo com o RGPD e que se encontra descrita em <http://fakeprofiles.online/privacy-policy.aspx>

O utilizador que pretender verificar se um determinado perfil do Twitter é considerado falso deve primeiramente recolher o *screen name* desse perfil. Para isso uma das formas é aceder à página desse perfil e copiar o *screen name* que se encontra no URL ou por baixo da imagem do perfil. No exemplo apresentado na Figura 14 o *screen name* é "realDonaldTrump" e o símbolo "@" não pertence ao *screen name*.



Figura 14 - Exemplo de uma página de perfil de um utilizador do Twitter.

Após a recolha do screen name deverá ir ao endereço da aplicação "fakeprofiles.online", aceitar os cookies e a política de privacidade, introduzir o *screen name* na caixa de texto e seleccionar o botão "SEARCH", tal como indicam os passos representados na Figura 15.

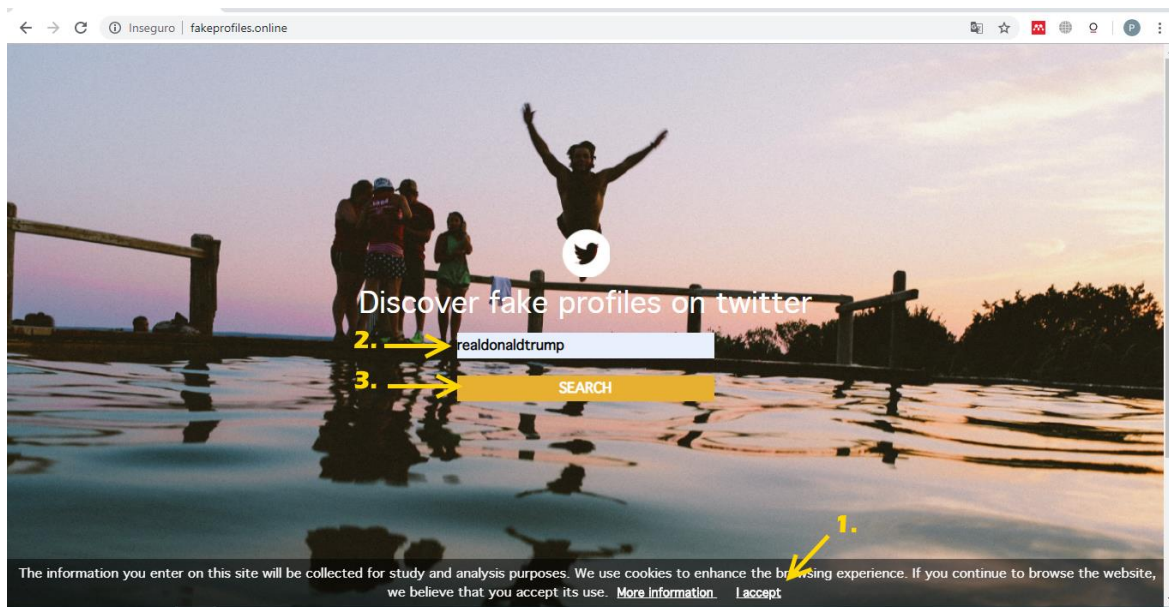


Figura 15 - Exemplo de utilização da aplicação.

A aplicação irá então questionar a API do Twitter. Quando obtiver as respostas analisará cada uma delas às quais atribuirá um valor, que após o cálculo com o peso, é então utilizado para o cálculo final, expresso em % de falsidade do perfil (Figura 16).

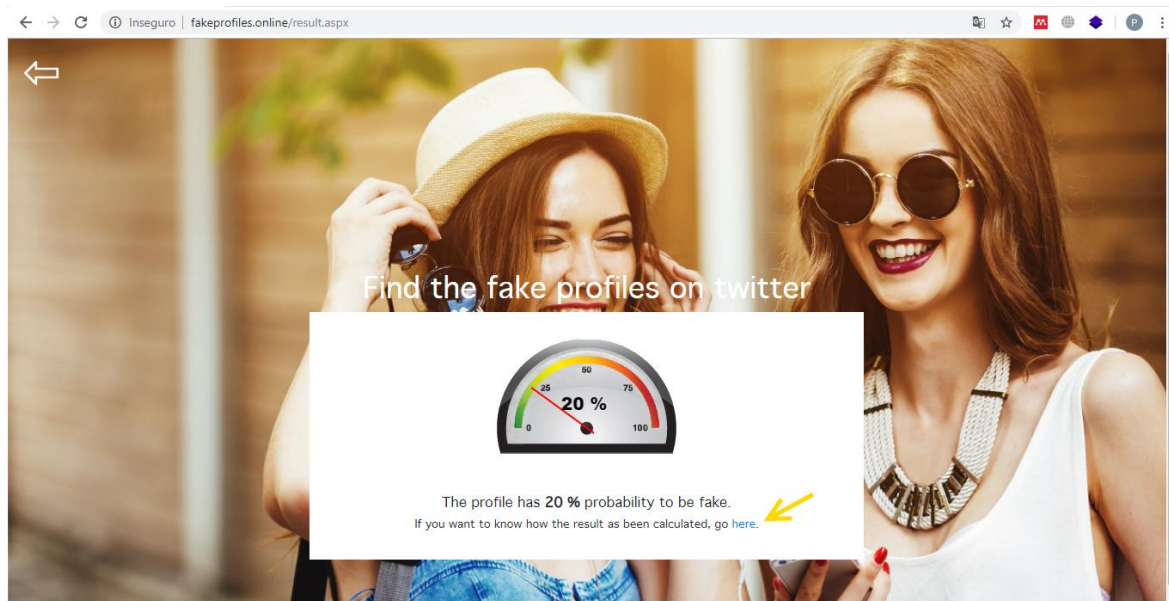
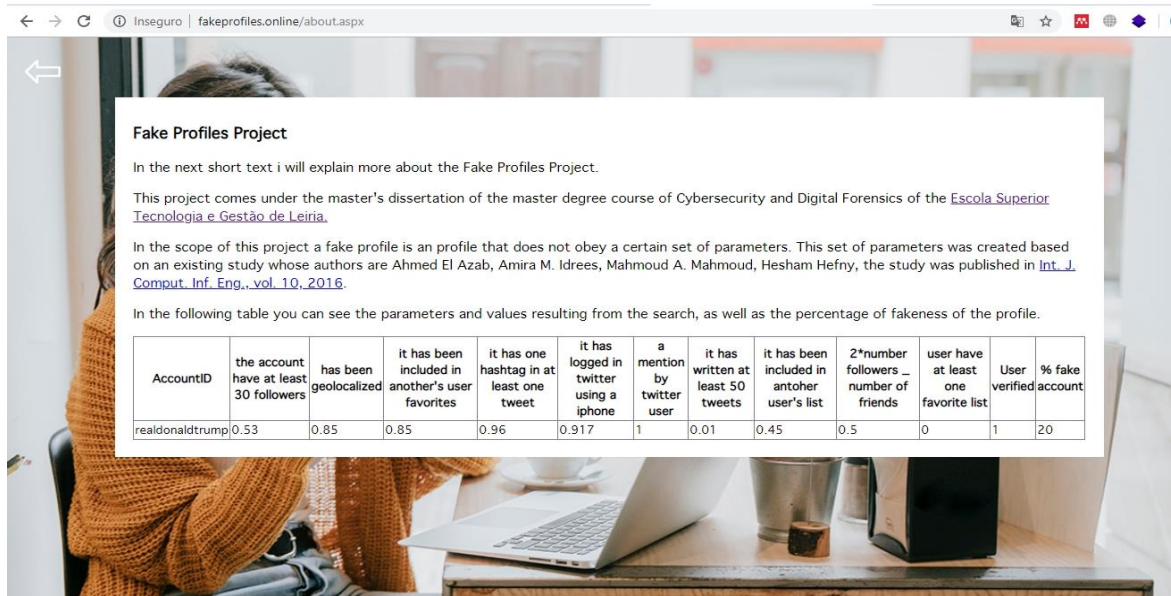


Figura 16 - Exemplo da página do resultado de uma pesquisa de um perfil.

Caso o utilizador pretenda saber mais informações, como quais são os parâmetros pesquisados e qual o resultado obtido, assim como outras informações, pode selecionar a opção “here” assinalada com uma seta amarela na Figura 16. Será então redirecionado para uma nova página onde são mostrados os detalhes do processamento efetuado, designadamente o resultado dos parâmetros questionados, o peso associado e o valor calculado. Poderá igualmente obter mais informações sobre o projeto. Na Figura 17 ilustra.se um exemplo do resultado da execução de uma pesquisa.



**Fake Profiles Project**

In the next short text i will explain more about the Fake Profiles Project.

This project comes under the master's dissertation of the master degree course of Cybersecurity and Digital Forensics of the [Escola Superior Tecnologia e Gestão de Leiria](#).

In the scope of this project a fake profile is a profile that does not obey a certain set of parameters. This set of parameters was created based on an existing study whose authors are Ahmed El Azab, Amira M. Idrees, Mahmoud A. Mahmoud, Hesham Hefny, the study was published in [Int. J. Comput. Inf. Eng., vol. 10, 2016](#).

In the following table you can see the parameters and values resulting from the search, as well as the percentage of fakeness of the profile.

AccountID	the account have at least 30 followers	has been geolocalized	it has been included in another's user favorites	it has one hashtag in at least one tweet	it has logged in twitter using a iphone	a mention by twitter user	it has written at least 50 tweets	it has been included in antoher user's list	2*number followers _ number of friends	user have at least one favorite list	User verified	% fake account
realDonaldTrump	0.53	0.85	0.85	0.96	0.917	1	0.01	0.45	0.5	0	1	20

**Figura 17 - Página com informação do projeto e detalhes dos parâmetros pesquisados.**

Resumindo e como já foi descrito, o projeto incidiu sobre a rede social Twitter. Esta opção deveu-se às limitações impostas atualmente pelas API das redes sociais Facebook e Instagram. Na implementação optei por desenvolver uma aplicação web, ficando assim disponível a todos os utilizadores da internet, utilizei a framework ASP.NET e recurso como linguagem C# pois é a que estou mais familiarizado.

Os parâmetros utilizados para identificação de um perfil falso foram baseados no estudo descrito em [55], sendo que adicionei o parâmetro “Conta estar validada”, por considerar que poderá fazer a diferença na identificação de perfis falsos.



## 4. Testes e Análise de Resultados

O presente capítulo descreve como foram criados os *dataset* utilizados nos testes e a sua origem, como foram efetuados os testes e a sua metodologia. Apresenta também a análise dos resultados obtidos.

### 4.1. *Dataset* de testes para a rede social Twitter

#### 4.1.1. Origem

Para a realização dos testes em relação à rede social Twitter foram construídos dois *datasets*, um para as contas reais (verdadeiras ou fidedignas) e outro para as contas falsas, ambos em formato .CSV, com os *screen name* dos utilizadores a pesquisar.

Os *dataset* utilizados resultam de um conjunto de *datasets* já utilizados pelo Projeto MIB [77], o qual os cedeu para eu poder utilizar. A informação dos *datasets* do projeto MIB encontra-se na Tabela 2.

Tabela 2 - Informação dos *dataset* do projeto MIB.

group name	Description	accounts	tweets	year
<b>genuine accounts</b>	verified accounts that are human-operated	3,474	8,377,522	2011
<b>social spambots #1</b>	retweeters of an Italian political candidate	991	1,610,176	2012
<b>social spambots #2</b>	spammers of paid apps for mobile devices	3,457	428,542	2014
<b>social spambots #3</b>	spammers of products on sale at Amazon.com	464	1,418,626	2011
<b>traditional spambots #1</b>	training set of spammers used by C. Yang, R. Harkreader, and G. Gu.	1,000	145,094	2009
<b>traditional spambots #2</b>	spammers of scam URLs	100	74,957	2014
<b>traditional spambots #3</b>	automated accounts spamming job offers	433	5,794,931	2013
<b>traditional spambots #4</b>	another group of automated accounts spamming job offers	1,128	133,311	2009
<b>fake followers</b>	simple accounts that inflate the number of followers of another account	3,351	196,027	2012

#### 4.1.2. Criação do *dataset*

Para as contas reais foi utilizado como base o *dataset* “*genuine accounts*”, com 3.474 de 2011. Para o primeiro teste utilizei uma parte do *dataset* e com a análise e validação verifiquei que muitas das contas presentes já não existiam atualmente. Para não enviesar o resultado dos testes era necessário ter um *dataset* só com contas reais e ativas, por isso foi feita uma limpeza ao *dataset* tendo sido selecionadas 100 contas existentes à data atual.

Da mesma forma foi necessário fazer essa limpeza para o *dataset* das contas falsas. Como base para a criação do *dataset* utilizei o *dataset* “*fake followers*” com 3.351 contas de 2012, do qual foram escolhidas 100 contas ativas para a criação do *dataset* de contas falsas.

## 4.2. Definição de testes e metodologia

Para agilizar e automatizar o teste fiz uma alteração na aplicação para que esta processasse automaticamente um ficheiro .CSV contendo os *screen names* dos utilizadores do *dataset*. O resultado do processamento das contas inseridas no ficheiro será apresentado para a análise aos 11 parâmetros.

Para elaboração do .CSV e dado as condicionantes já mencionadas na secção 3.3, criei 15 ficheiros .CSV, 6 deles com 15 screennames e um outro com 10, perfazendo assim os 100 utilizadores presentes no *dataset* de contas verdadeiras. Para as contas falsas foram criados o mesmo número de ficheiros .CSV.

O teste aos dois *datasets* decorreu no mesmo dia e foram elaborados durante o mês de março. Após realizados os testes criei um ficheiro .xlsx com a agregação dos resultados obtidos para as contas verdadeiras e outro para as contas falsas. Para efeitos de teste foram considerados os 11 parâmetros e feito depois várias aproximações para 10 parâmetros, 7 parâmetros e 8 parâmetros. Seguidamente foi feita a análise de resultados.

## 4.3. Resultados obtidos

Os valores de cada parâmetro e a percentagem de falsidade obtida nos testes efetuados com o *dataset* encontram-se na sua totalidade no Anexo A e B. Dos dados obtidos e após tratamento estatístico dos mesmo extraí a informação presente nas seguintes tabelas.

A Tabela 3 representa a média da “percentagem de falsidade” para cada conjunto de parâmetros e para cada um dos dois *dataset*: contas reais e contas falsas.

Tabela 3 - Média da percentagem de falsidade

	11 parâmetros	10 parâmetros	8 parâmetros	7 parâmetros
<b>Dataset contas Reais</b>	52,92	55,88	41,38	33,59
<b>Dataset contas Falsas</b>	87,73	86,50	83,13	80,71

A Tabela 5 representa o número de contas por intervalo de percentagem de falsidade para o *dataset* das contas falsas.

Tabela 4 - Número de contas por percentagem de falsidade, *dataset* contas reais.

Valor	>=40 %	>=50 %	>=60 %	>=70 %	>=80 %	>=90 %
<b>11 parâmetros</b>	96	66	41	24	15	9
<b>10 parâmetros</b>	70	45	29	22	15	9

<b>8 parâmetros</b>	64	41	29	19	11	7
<b>7 parâmetros</b>	45	29	20	19	10	7

A Tabela 5 representa o número de contas por intervalo de percentagem de falsidade para o *dataset* das contas falsas.

**Tabela 5 - Número de contas por % de falsidade, *dataset* contas falsas.**

Valor	>=40 %	>=50 %	>=60 %	>=70 %	>=80 %	>=85 %	>=90 %	>=95 %
<b>11 parâmetros</b>	100	100	100	100	99	94	40	27
<b>10 parâmetros</b>	100	100	100	100	97	91	40	27
<b>8 parâmetros</b>	100	100	100	100	97	45	33	3
<b>7 parâmetros</b>	100	100	100	97	96	45	33	3

#### 4.4. Análise de resultados

Analisando a Tabela 3, em relação ao *dataset* de contas reais e utilizando o conjunto dos 11 parâmetros temos uma média de 52,92%, para o conjunto de 10 parâmetros 55,88%, 8 parâmetros 41,38% e para o conjunto de 7 parâmetros 33,59%. Para o *dataset* de contas falsas temos uma média de 87,73% para o conjunto de 11 parâmetros, 86,5% para 10 parâmetros, 83,13% para o conjunto de 8 parâmetros e 80,71% para o de 7.

O gráfico da Figura 18 representa a distribuição da média de falsidade para os *dataset* das contas verdadeiras e falsas em relação ao conjunto de 11, 10, 8 e 7 parâmetros. A linha azul representa o *dataset* de contas reais e a linha vermelha representa o *dataset* de contas falsas.

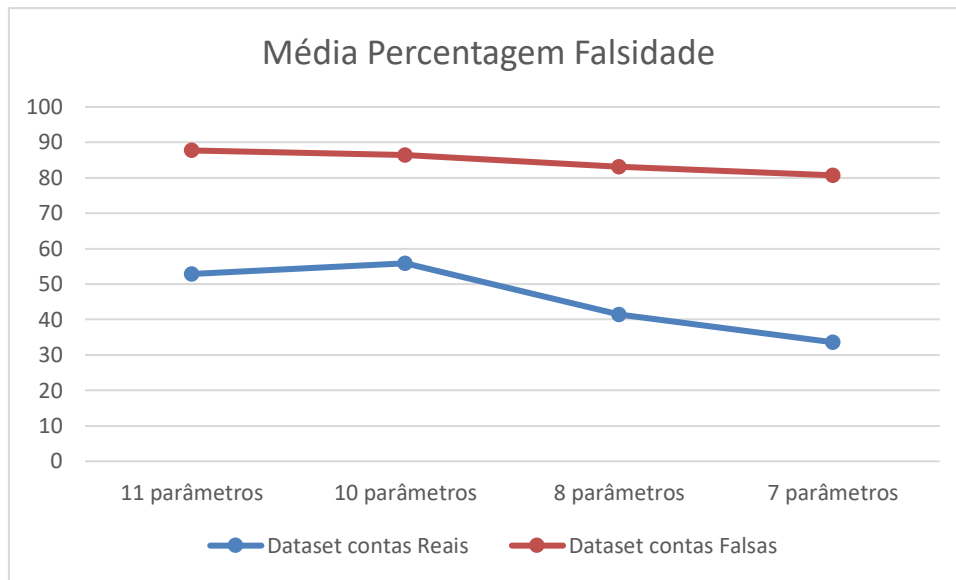


Figura 18 - Média percentagem falsidade.

A Tabela 4 representa o número de contas por intervalo de percentagem de falsidade para o *dataset* das contas reais. Analisando a tabela para o caso de 11 parâmetros podemos verificar que, se considerarmos um *threshold* de 40%, temos 96 contas classificadas como falsas, com valor maior ou igual a 50% temos 66 contas consideradas falsas, com valor igual ou superior a 60% ficaríamos com 41 contas classificadas como falsas.

No entanto, com valor superior ou igual a 70% o número de contas falsas reduz para 24, com uma percentagem igual ou superior a 80% reduz para 15 contas e com um valor igual ou superior a 90% só classificou 9 contas como falsas. O intervalo de valores onde se regista maior diferença é entre os 60% e os 70%.

Para o conjunto dos 10 parâmetros e ainda tendo como referência a Tabela 4 e com um *threshold* de 40% obtiveram-se 70 contas classificadas como falsas, com um valor igual ou superior a 50% diminuiu para 45, sendo que com um valor igual ou superior a 60% registou-se 29, com valor superior ou igual a 70% 22, com 80% apenas 15 e a partir 90% só 9 foram classificadas como falsas. Neste caso o intervalo onde se registou um maior decréscimo foi entre os 40% e 50%.

No caso do conjunto de 8 parâmetros registou-se apenas 64 contas classificadas como falsas com um *threshold* de 40%, verifica-se logo aqui um menor número de contas classificadas como falsas em relação ao mesmo valor de *threshold* para o conjunto de 11 e 10 parâmetros, ou seja, a metodologia utilizando apenas 8 parâmetros é melhor que utilizando 11 ou 10.

Para um valor igual ou superior a 50% registou-se 41 contas classificadas como falsas, sendo que com um valor superior ou igual a 60% o número é de 29 reduzindo para 19 caso o valor seja igual ou maior a 70%, com valores superiores a 80% e 90% já se registou apenas 11 e 7 respetivamente. No caso do conjunto com 8 parâmetros a maior diferença registou-se entre os 40% e os 50%, também aqui e em comparação com o conjunto de 11 e 10 parâmetros esta diferença foi registada antes.

Analisando a Tabela 4 para o conjunto de 7 parâmetros registou-se desde logo um menor número de contas classificadas como falsas quando comparado com os conjuntos de parâmetros anteriormente apresentados e para os mesmos intervalos.

O número de contas classificadas como falsas chega a ser menos de metade em comparação o conjunto de 11 parâmetros para valores entre os 40% e os 60%. Registaram-se cerca de 45 contas consideradas falsas para valores de threshold de 40%, sendo que para valores de 50% o número foi de 29, para valores de 60% foram 20 e para valores de 70%, 80% e 90% foram 19, 10 e 7 respetivamente. Neste conjunto de parâmetros a maior diferença registou-se entre os 40% e os 50%.

Comparando o conjunto de 8 parâmetros com o de 7 parâmetros e para um threshold de 90% verifica-se que o número de contas classificadas é o mesmo, mas nos restantes valores de threshold verifica-se que o número de contas falsas é inferior no conjunto de 7 parâmetros quando comparado com o conjunto de 8 parâmetros. Garantindo assim que a melhor aproximação é utilizando o conjunto de 7 parâmetros, pois é a que nos oferece melhores resultados.

Tomando como base os dados da Tabela 4 e o descrito anteriormente, encontram-se representado na Tabela 6 os falsos positivos para um valor de percentagem de falsidade superior a 80% e 90% para o *dataset* de contas reais. Se considerarmos como conta falsa uma conta que obtenha um valor de percentagem de falsidade superior a 80% temos então para o *dataset* de contas reais 10 contas classificadas como falsas, ou seja, são 10 falsos negativos. Como o *dataset* tinha 100 contas atingimos então uma percentagem de assertividade de 90%.

Se aumentar o threshold para 90% ficamos apenas com 7 falsos negativos e conseguimos neste caso 93% de assertividade tal como se pode verificar na Tabela 7 que representa a percentagem de assertividade para o conjunto de 7 parâmetros do *dataset* das contas reais com threshold superior a 80% e 90%.

**Tabela 6 - Falsos negativos, *dataset* contas reais, conjunto 7 parâmetros.**

Threshold	>=80 %	>=90 %
FN	10	7

**Tabela 7 - Percentagem Assertividade, *dataset* contas reais, conjunto 7 parâmetros.**

Threshold	>=80 %	>=90 %
% assertividade	90%	93%

A Tabela 5 representa o número de contas por intervalo de percentagem de falsidade para o *dataset* das contas falsas. Verifica-se que até ao valor de 60% todas as contas do *dataset* são classificadas como falsas, para qualquer conjunto de parâmetros. Para um threshold de 70% todas as contas são classificadas como falsas, à exceção do conjunto de 7 parâmetros que apenas classificou como falsas 97 contas.

No conjunto com 11 parâmetros e para um threshold de 80% e 85% registou-se 99 e 94 contas classificadas como falsas respetivamente, sendo que para valores de 90% e 95% o número de contas diminui para 40 e 27. Para o conjunto de 10 parâmetros e com threshold de 80% e 85% o número de contas classificadas como falsas é de 97 e 91 apresentando uma diminuição drástica para valores de threshold de 90% e 95% cujo número de contas classificadas como falsas é de 40 e 27 respetivamente.

Para o conjunto de 8 e 7 parâmetros e com threshold de 85%, 90% e 95% temos o mesmo número de contas classificadas como falsas, ou seja, 45,33 e 3. Com o threshold de 80% apenas difere em uma conta sendo que com 8 parâmetros foram classificadas como falsas 97 e com 7 parâmetros 96.

Em relação ao intervalo onde se registou a maior diferença, analisando a Tabela 5 pode verificar-se que para o conjunto de 11 e 10 parâmetros foi entre os 85% e os 90%, sendo que no caso do conjunto dos 8 e 7 parâmetros foi entre os 90% e 95%.

Em análise à Tabela 5 podemos concluir que para o caso do *dataset* de contas falsas a melhor abordagem é o conjunto de 11 parâmetros, pois é o que ao longo do aumento do threshold nos indica sempre um maior número de contas falsas em comparação com os restantes.

Tendo em conta a melhor aproximação para o *dataset* das contas falsas, está representado na Tabela 8 os falsos positivos, ou seja as contas que são consideradas como verdadeiras, então temos para um valor igual ou superior a 80% temos 99 contas consideradas falsas e apenas 1 considerada verdadeira, se diminuirmos o intervalo e considerarmos os 85% ficamos com 94 consideradas falsas e 6 verdadeiras, ou seja, 6 falsas positivas.

**Tabela 8 - Falsos positivo, *dataset* contas falsas, conjunto 11 parâmetros.**

Threshold	>=80 %	>=85 %
FP	1	6

**Tabela 9 - Percentagem Assertividade, *dataset* contas falsas, conjunto 11 parâmetros.**

Threshold	>=80 %	>=85 %
% assertividade	99%	94%

Tendo em conta os dados representados na Tabela 4 e para valores iguais ou superiores 80% e na Tabela 5 para valores iguais ou superiores a 90%, comparando os valores dos conjuntos de parâmetros 7 com 8 e dos parâmetros 10 com 11 podemos verificar que o parâmetro adicionado “Conta estar validada” não têm impacto nos valores para os intervalos referidos.

## 5. Conclusões e Trabalho Futuro

Torna-se cada vez mais importante dotar as pessoas e as empresas de conhecimento, ferramentas e informação acerca dos perigos que podem correr na Internet.

Nos últimos anos e como já foi descrito neste documento o aumento de redes sociais tem sido desmesurado. Com esse aumento surgiram cada vez mais problemas de segurança e aumento do cibercrime devido à existência de perfis falsos nas redes sociais. O problema neste momento e a nível transversal para todas as redes sociais, consiste em conseguir identificar, de forma eficaz e simples, perfis falsos.

Pretendi com este trabalho deixar assim o meu contributo para este problema, através da definição de uma metodologia de apoio à decisão sobre a veracidade de um perfil da rede social Twitter. A contribuição estendeu-se também ao desenvolvimento de uma aplicação para o utilizador comum da Internet, que aplica a metodologia e o algoritmo definidos neste trabalho, embora baseados em trabalho já realizado por outros autores.

A aplicação web que se encontra disponível que por base o *screen name* de um utilizador da rede social Twitter e faz uma pesquisa recorrendo à API da rede social. Com base num conjunto de parâmetros classifica depois esse perfil em conta verdadeira ou falsa. Não foi possível alargar este desenvolvimento ao Facebook e Instagram devido a restrições impostas pelos mesmos na sua API.

Foram elaborados testes que permitiram demonstrar a validade e viabilidade da solução apresentada para o problema, a metodologia implementada permitiu atingir um nível de assertividade entre os 90% e 99%.

Em termos de trabalho futuro, existem várias linhas de desenvolvimento que podem ser seguidas. Uma dessas linhas consiste em verificar se com recurso a API pagas das redes sociais seria possível obter a informação que não foi possível obter neste projeto, ou verificar se seria possível recolher a informação de alguma outra forma.

Outra linha de desenvolvimento sugerida, seria a criação de uma aplicação móvel com uma base de dados centralizada.

Outro ponto interessante seria deixar o utilizador, aquando da pesquisa, seleccionar um espectro mais alargado de parâmetros. Também a recolha da informação e disponibilização desta ao utilizador poderia ser uma mais valia.

## Bibliografia ou Referências Bibliográficas

- [1] Mariana Oliveira, “Criação de falsos perfis nas redes sociais para difamar no topo das queixas de cibercrime | Justiça | PÚBLICO,” 2014. [Online]. Available: <https://www.publico.pt/2014/03/30/sociedade/noticia/criacao-de-falsos-perfis-nas-redes-sociais-para-difamar-no-topo-das-queixas-de-cibercrime-1630283>. [Accessed: 10-Nov-2018].
- [2] JOÃO PORFÍRIO, “Cinco pessoas detidas pela PJ após crimes de roubo violentos com recurso a perfis falsos no Facebook – Observador,” 2019. [Online]. Available: <https://observador.pt/2019/05/06/cinco-pessoas-detidas-pela-pj-apos-crimes-de-roubo-violentos-com-recurso-a-perfis-falsos-no-facebook/>. [Accessed: 01-Aug-2019].
- [3] “Detido ex-professor suspeito de criar perfis falsos para aliciar menores - Atualidade - SAPO 24,” 2019. [Online]. Available: <https://24.sapo.pt/atualidade/artigos/detido-ex-professor-suspeito-de-criar-perfis-falsos-para-aliciar-menores>. [Accessed: 01-Aug-2019].
- [4] NELSON MORAIS, “Criação de perfil falso no Facebook não foi considerada crime,” 2016. [Online]. Available: <https://www.jn.pt/justica/interior/criacao-de-perfil-falso-no-facebook-nao-foi-considerada-crime-5520453.html>. [Accessed: 01-Aug-2019].
- [5] A. El Azab, A. M. Idrees, M. A. Mahmoud, and H. Hefny, “Fake Account Detection in Twitter Based on Minimum Weighted Feature set,” Nov. 2015.
- [6] danah m. boyd and N. B. Ellison, “Social Network Sites: Definition, History, and Scholarship,” *J. Comput. Commun.*, vol. 13, no. 1, pp. 210–230, Oct. 2007.
- [7] jamie, “65+ Social Networking Sites You Need to Know About in 2019 - Make A Website Hub.” [Online]. Available: <https://makeawebsitehub.com/social-media-sites/>. [Accessed: 08-Aug-2019].
- [8] S. Phillips, “A brief history of Facebook,” *Guard.*, Jul. 2007.
- [9] “And Facebook begins,” *The Sunday Indian*, May 2012.
- [10] L. O’Brien, “Poking Facebook,” *:02138*, p. 66, 2007.
- [11] B. Schwartz, “Hot or Not? Website Briefly Judges Looks,” *Harvard Crimson*, Nov. 2003.
- [12] M. Bellis, “Who Invented Facebook?,” *about.com*.



- [13] Salman Aslam, “• Facebook by the Numbers (2018): Stats, Demographics & Fun Facts,” 2018. [Online]. Available: <https://www.omnicoreagency.com/facebook-statistics/>. [Accessed: 10-Oct-2018].
- [14] “Our Story – Instagram.” [Online]. Available: <https://instagram-press.com/our-story/>. [Accessed: 08-Oct-2018].
- [15] E. M. Rusli, “Facebook Buys Instagram for \$1 Billion - The New York Times,” *April 9, 2012 1:15 pm*, 2012. [Online]. Available: <https://dealbook.nytimes.com/2012/04/09/facebook-buys-instagram-for-1-billion/>. [Accessed: 08-Oct-2018].
- [16] Salman Aslam, “• Instagram by the Numbers (2018): Stats, Demographics & Fun Facts,” 2018. [Online]. Available: <https://www.omnicoreagency.com/instagram-statistics/>. [Accessed: 10-Oct-2018].
- [17] “Fenix 2 for Twitter – Aplicações no Google Play.” [Online]. Available: <https://play.google.com/store/apps/details?id=it.mvilla.android.fenix2>. [Accessed: 17-Sep-2019].
- [18] “Friendly For Twitter - Apps on Google Play.” [Online]. Available: [https://play.google.com/store/apps/details?id=io.friendly.twitter&hl=en\\_US](https://play.google.com/store/apps/details?id=io.friendly.twitter&hl=en_US). [Accessed: 17-Sep-2019].
- [19] “The Real History of Twitter, In Brief.” [Online]. Available: <https://www.lifewire.com/history-of-twitter-3288854>. [Accessed: 13-Jan-2019].
- [20] Salman Aslam, “• Twitter by the Numbers (2018): Stats, Demographics & Fun Facts,” *September 18, 2018*, 2018. [Online]. Available: <https://www.omnicoreagency.com/twitter-statistics/>. [Accessed: 10-Oct-2018].
- [21] “População mundial atingiu 7,6 bilhões de habitantes | ONU News.” [Online]. Available: <https://news.un.org/pt/story/2017/06/1589091-populacao-mundial-atingiu-76-bilhoes-de-habitantes>. [Accessed: 10-Oct-2018].
- [22] “Products Overview — Twitter Developers.” [Online]. Available: <https://developer.twitter.com/en/products/products-overview>. [Accessed: 08-Jan-2019].
- [23] “curl for Windows.” [Online]. Available: <https://curl.haxx.se/windows/>. [Accessed: 19-Mar-2019].
- [24] “Instagram Developer Documentation.” [Online]. Available: <https://www.instagram.com/developer/>. [Accessed: 19-Mar-2019].
- [25] “Proteção de Dados - RGPD ( Regulamento Geral Proteção de Dados ).” [Online].

- Available: <https://protecao-dados.pt/o-regulamento/>. [Accessed: 18-Mar-2019].
- [26] N. C. Matthew Rosenberg and C. Cadwalladr, *How Trump Consultants Exploited the Facebook Data of Millions*. The New York Times, 2018.
- [27] Kaleigh Rogers, “Facebook Posts ‘Substantively Contributed’ to Myanmar Genocide, UN Investigators Say - VICE.” [Online]. Available: [https://www.vice.com/en\\_us/article/7x7gmx/facebook-un-genocide-rohingya-myanmar-hate-speech](https://www.vice.com/en_us/article/7x7gmx/facebook-un-genocide-rohingya-myanmar-hate-speech). [Accessed: 25-Jun-2019].
- [28] S. Perez, “Facebook rolls out more API restrictions and shutdowns | TechCrunch,” 2018. [Online]. Available: <https://techcrunch.com/2018/07/02/facebook-rolls-out-more-api-restrictions-and-shutdowns/>. [Accessed: 02-Sep-2019].
- [29] “Rate Limiting — Twitter Developers.” [Online]. Available: <https://developer.twitter.com/en/docs/basics/rate-limiting.html>. [Accessed: 25-Jun-2019].
- [30] “Rate Limits - Graph API.” [Online]. Available: <https://developers.facebook.com/docs/graph-api/overview/rate-limiting#application-level-rate-limiting>. [Accessed: 25-Jun-2019].
- [31] “Instagram Platform API for Developers | Facebook for Developers.” [Online]. Available: <https://developers.facebook.com/products/instagram/>. [Accessed: 02-Jul-2019].
- [32] Joseph Cox, “Facebook mudou discretamente uma ferramenta de busca usada por investigadores, mas abusada por empresas - VICE,” 2019. [Online]. Available: [https://www.vice.com/pt\\_br/article/zmpgmx/facebook-mudou-discretamente-uma-ferramenta-de-busca-usada-por-investigadores-mas-abusada-por-empresas](https://www.vice.com/pt_br/article/zmpgmx/facebook-mudou-discretamente-uma-ferramenta-de-busca-usada-por-investigadores-mas-abusada-por-empresas). [Accessed: 25-Jun-2019].
- [33] L. Brown, “Beware of the bots: How they’re created and why they matter - TechRepublic,” 2017. [Online]. Available: <https://www.techrepublic.com/article/beware-of-the-bots-how-theyre-created-and-why-they-matter/>. [Accessed: 20-Sep-2019].
- [34] R. H. and M. H. NICHOLAS CONFESSORE, GABRIEL J.X. DANCE, “The Follower Factory - The New York Times.” [Online]. Available: <https://www.nytimes.com/interactive/2018/01/27/technology/100000005704904.app.html>. [Accessed: 19-Nov-2018].
- [35] Facebook, “Facebook Publishes Enforcement Numbers for the First Time | Facebook Newsroom,” 2018. [Online]. Available:

- <https://newsroom.fb.com/news/2018/05/enforcement-numbers/>. [Accessed: 23-May-2018].
- [36] “Consulte o significado / definição de cibercrime no Dicionário Priberam da Língua Portuguesa, o dicionário online de português contemporâneo.” [Online]. Available: <https://dicionario.priberam.org/cibercrime>. [Accessed: 10-Nov-2018].
- [37] “Social media-related crime reports up 780% in four years | Media | The Guardian.” [Online]. Available: <https://www.theguardian.com/media/2012/dec/27/social-media-crime-facebook-twitter>. [Accessed: 15-Oct-2018].
- [38] Annabelle Graham, “The rise of cyber crime - IT Governance Blog,” 2018. [Online]. Available: <https://www.itgovernance.co.uk/blog/the-rise-of-cyber-crime>. [Accessed: 10-Nov-2018].
- [39] “Six types of killer use Facebook to commit crimes, says study | Technology | The Guardian.” [Online]. Available: <https://www.theguardian.com/technology/2014/nov/03/killer-facebook-social-media-violent-criminologists>. [Accessed: 10-Nov-2018].
- [40] Direcção Nacional, “Polícia de Segurança Pública :: Noticias :: Detalhe.” [Online]. Available: <http://www.psp.pt/Pages/Noticias/MostraNoticia.aspx?NoticiasID=728>. [Accessed: 06-Nov-2018].
- [41] Polícia Segurança Pública, “Será mesmo necessário publicar fotos com... - Polícia Segurança Pública | Facebook.” [Online]. Available: <https://www.facebook.com/policiasegurancapublica/photos/pb.109274852461371.-2207520000.1440439018./867133950008787/?type=3&theater>. [Accessed: 08-Nov-2018].
- [42] G. Cibecrime, “relatorio da atividade\_cibercrime\_2013.”
- [43] Setembro and Dezembro, “Gabinete Cibercrime RELATÓRIO DA ACTIVIDADE SETEMBRO 2015 a DEZEMBRO 2016.”
- [44] Nuno Ribeiro Coelho, “Acórdão do Tribunal da Relação do Porto.” [Online]. Available: <http://www.dgsi.pt/jtrp.nsf/56a6e7121657f91e80257cda00381fdf/3e55763d907f780280257f6b004d4ce6?OpenDocument&Highlight=0,Facebook>. [Accessed: 10-Nov-2018].
- [45] Fátima Furtado, “Acórdão do Tribunal da Relação do Porto.” [Online]. Available: <http://www.dgsi.pt/jtrp.nsf/d1d5ce625d24df5380257583004ee7d7/872f3063233d8de480257b78003e60f3?OpenDocument>. [Accessed: 10-Nov-2018].

- [46] Pedro Sales Dias, “PJ alerta para aumento de crimes com ‘sextortion’ nas redes sociais | Internet | PÚBLICO.” [Online]. Available: <https://www.publico.pt/2015/09/11/sociedade/noticia/pj-alerta-para-aumento-de-crimes-com-sextortion-nas-redes-sociais-1707521>. [Accessed: 10-Nov-2018].
- [47] lusa, “PJ de Leiria recebe entre duas a quatro queixas por dia de crimes informáticos.” [Online]. Available: <https://www.dn.pt/lusa/interior/pj-de-leiria-recebe-entre-duas-a-quatro-queixas-por-dia-de-crimes-informaticos-8486758.html>. [Accessed: 12-Nov-2018].
- [48] E. Graham-Harrison and C. Cadwalladr, “Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach,” *the Guardian*, 2018. [Online]. Available: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>. [Accessed: 14-May-2019].
- [49] Paulo Pena, “Redes sociais: há um consultor do PSD numa campanha de perfis falsos,” *DN*, 2019. [Online]. Available: <https://www.dn.pt/poder/interior/redes-sociais-ha-um-consultor-do-psd-numa-campanha-de-perfis-falsos--10887064.html>. [Accessed: 14-May-2019].
- [50] “PS apresenta queixa à CNE por campanha de perfis falsos nas redes sociais.” [Online]. Available: <https://www.jn.pt/nacional/canal/europeias-2019/interior/ps-apresenta-queixa-a-cne-por-campanha-de-perfis-falsos-nas-redes-sociais-10897090.html>. [Accessed: 14-May-2019].
- [51] M. Conti, R. Poovendran, and M. Secchiero, “FakeBook: Detecting Fake Profiles in On-Line Social Networks,” in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2012, pp. 1071–1078.
- [52] A. L. Zifei Shan, Haowen Cao, Jason Lv, Cong Yan, “Enhancing and Identifying Cloning Attacks in Online Social Networks,” 2013.
- [53] S. Gurajala, J. S. White, B. Hudson, and J. N. Matthews, “Fake Twitter accounts: Profile characteristics obtained using an activity-based pattern detection approach,” 2015.
- [54] D. Kagan, Y. Elovichi, and M. Fire, “Generic anomalous vertices detection utilizing a link prediction algorithm,” *Soc. Netw. Anal. Min.*, vol. 8, no. 1, p. 27, Dec. 2018.
- [55] H. H. Ahmed El Azab, Amira M. Idrees, Mahmoud A. Mahmoud, “Fake Account Detection in Twitter Based on Minimum Weighted Feature set,” *Int. J. Comput. Inf. Eng.*, vol. 10, 2016.
- [56] L. Bilge, T. Strufe, D. Balzarotti, E. Kirda, and S. Madrid, *All Your Contacts Are*

- Belong to Us: Automated Identity Theft Attacks on Social Networks*. 2009.
- [57] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, *The Socialbot Network: When Bots Socialize for Fame and Money*. 2011.
- [58] C. Wagner, S. Mitter, C. Körner, and M. Strohmaier, “When social bots attack: Modeling susceptibility of users in online social networks.”
- [59] T. Rodrigues, “Detecting Spammers on Twitter.” .
- [60] X. Zheng, Z. Zeng, Z. Chen, Y. Yu, and C. Rong, “Detecting spammers on social networks,” *Neurocomputing*, vol. 159, pp. 27–34, Jul. 2015.
- [61] “The Fake Project.” [Online]. Available: <http://wafi.iit.cnr.it/fake/fake/app/>. [Accessed: 02-Jan-2019].
- [62] L. Breiman, *RANDOM FORESTS*. 2001.
- [63] O. Maimon and L. Rokach, Eds., *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US, 2010.
- [64] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, “An Introduction to Information Retrieval,” 2009.
- [65] D. Kriesel, “Neural Networks.”
- [66] T. Joachims, “Text Categorization with Support Vector Machines: Learning with Many Relevant Features.”
- [67] A. Gowda Karegowda, A. S. Manjunath, and M. A. Jayaram, “COMPARATIVE STUDY OF ATTRIBUTE SELECTION USING GAIN RATIO AND CORRELATION BASED FEATURE SELECTION.”
- [68] T. Mori, M. Kikuchi, and K. Yoshida, “Term Weighting Method based on Information Gain Ratio for Summarizing Documents retrieved by IR systems.”
- [69] J. Davis and M. Goadrich, “The Relationship Between Precision-Recall and ROC Curves.”
- [70] R. dos S. Leal, “Métricas Comuns em Machine Learning: como analisar a qualidade de chat bots inteligentes — métricas (3 de 4).” [Online]. Available: <https://medium.com/as-máquinas-que-pensam/métricas-comuns-em-machine-learning-como-analisar-a-qualidade-de-chat-bots-inteligentes-métricas-1ba580d7cc96>. [Accessed: 22-Sep-2019].
- [71] Michael Bazzell, “IntelTechniques.com | OSINT & Privacy Services by Michael Bazzell | Open Source Intelligence.” [Online]. Available: <https://inteltechniques.com/>. [Accessed: 05-Jan-2019].
- [72] “IntelTechniques.com | OSINT & Privacy Services by Michael Bazzell | Open

- Source Intelligence.” [Online]. Available:  
<https://inteltechniques.com/osint/instagram.html>. [Accessed: 05-Jan-2019].
- [73] “Twitter Search Tool by IntelTechniques.” [Online]. Available:  
<https://inteltechniques.com/osint/twitter.html>. [Accessed: 05-Jan-2019].
- [74] “Facebook Search Tool by IntelTechniques.com.” [Online]. Available:  
<https://inteltechniques.com/osint/facebook.html>. [Accessed: 05-Jan-2019].
- [75] M. Bazzell, “125-What Happened? by The Privacy, Security, & OSINT Show.” [Online]. Available: <https://soundcloud.com/user-98066669/126-what-happened>. [Accessed: 19-Aug-2019].
- [76] T. Deusdedith, “Uma maneira infalível de ser verificado no Twitter | Agorapulse,” 2016. [Online]. Available: <https://www.agorapulse.com/pt/blog/uma-maneira-de-ser-verificado-no-twitter>. [Accessed: 03-Sep-2019].
- [77] M. P. Fazzolari, “My Information Bubble project.” [Online]. Available:  
<http://mib.projects.iit.cnr.it/index.html>. [Accessed: 15-Apr-2019].

## Anexos

### Anexo A - Resultados *dataset* contas reais

AccountID	the account have at least 30 followers	has been geolocalized	it has been included in another's user favorites	it has one hash tag in at least one tweet	it has logged in twitter using a iph one	a ment ion by twitt er user	it has writ ten at least 50 tweets	it has been inclu ded in anto her user' s list	2*nu mber follow ers _ numb er of friend s	user have at least one favo rite list	User verif ied	% fake account (10 param etros exclui user verifie d)	% fake account (7 param etros)	% fake account (11 param etros)	% fake account (8 param etros)
1120Roll	0,53	0,85	0,85	0,96	0,917	0	0,01	0,45	0	0,17	0	52,63	41,33	56,94	48,66
wadespete rs	0	0	0	0	0,917	0	0	0	0	0	0	90,83	86,90	91,66	88,54
191a5bd05 da04dc	0	0	0,85	0,96	0	0	0,01	0	0	0	0	81,8	74,14	83,45	77,38
19_Joanne _87	0,53	0	0,85	0,96	0,917	1	0,01	0,45	0,5	0	0	47,83	32,04	52,57	40,54
1Dniallprin cess	0,53	0,85	0,85	0,96	0,917	0	0,01	0,45	0,5	0,17	0	47,63	34,19	52,39	42,41
1GisellePiz arro	0,53	0,85	0,85	0,96	0	1	0,01	0,45	0,5	0	0	48,5	33,00	53,18	41,38

<b>1Nicoleromany</b>	0,53	0,85	0,85	0,96	0,917	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>1_DErika</b>	0,53	0,85	0,85	0,96	0,917	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>2cdevelopment</b>	0,53	0	0,85	0,96	0,917	0	0,01	0,45	0,5	0,17	0	56,13	46,33	60,12	53,04
<b>2hip4tv</b>	0,53	0	0,85	0,96	0	1	0,01	0,45	0,5	0	1	57	45,14	51,82	39,50
<b>510Daniel</b>	0,53	0,85	0,85	0,96	0,917	0	0,01	0,45	0	0	0	54,33	41,33	58,48	48,66
<b>AfnanAK</b>	0,53	0	0,85	0,96	0,917	0	0,01	0,45	0,5	0	0	57,83	46,33	61,66	53,04
<b>AgyapasJordan</b>	0,53	0	0,85	0,96	0	0	0,01	0,45	0,5	0	0	67	59,43	70,00	64,50
<b>AirJordan136</b>	0,53	0	0,85	0,96	0,917	0	0,01	0,45	0,5	0	0	57,83	46,33	61,66	53,04
<b>AkheemV</b>	0,53	0,85	0,85	0,96	0	1	0,01	0,45	0,5	0	0	48,5	33,00	53,18	41,38
<b>Akhtarul_Iman</b>	0,53	0	0,85	0,96	0	0	0,01	0	0	0	0	76,5	66,57	78,64	70,75
<b>AI_FSU</b>	0,53	0,85	0,85	0,96	0,917	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>AlaskaAmy</b>	0,53	0	0,85	0,96	0,917	0	0,01	0,45	0,5	0	0	57,83	46,33	61,66	53,04
<b>AlbertoDefa</b>	0,53	0,85	0,85	0,96	0,917	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>Albertopurple</b>	0,53	0,85	0,85	0,96	0,917	0	0,01	0,45	0,5	0	0	49,33	34,19	53,94	42,41
<b>AleksandrValeev</b>	0,53	0,85	0,85	0,96	0,917	0	0,01	0,45	0,5	0	0	49,33	34,19	53,94	42,41



<b>AlexBevan nn</b>	0,53	0,85	0,85	0,96	0,91 7	0	0,01	0,45	0,5	0	0	49,33	34,19	53,94	42,41
<b>Alex_Senior1</b>	0,53	0,85	0,85	0,96	0,91 7	0	0,01	0,45	0,5	0	0	49,33	34,19	53,94	42,41
<b>Alivnna</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>Juanma</b>	0,53	0	0,85	0,96	0	1	0,01	0,45	0,5	0	0	57	45,14	60,91	52,00
<b>Sam</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>andeh</b>	0	0	0	0,96	0	0	0	0	0,5	0	0	85,4	79,14	86,73	81,75
<b>Johnny</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>Jpinzon</b>	0,53	0	0,85	0,96	0,91 7	0	0,01	0,45	0,5	0	0	57,83	46,33	61,66	53,04
<b>Piglet</b>	0,53	0	0,85	0,96	0	0	0	0,45	0,5	0	0	67,1	59,43	70,09	64,50
<b>Yoyin</b>	0	0	0	0	0	0	0	0	0	0	0	100	100,00	100,00	100,00
<b>Joshua</b>	0,53	0	0,85	0,96	0	1	0,01	0,45	0,5	0	1	57	45,14	51,82	39,50
<b>julia</b>	0,53	0	0,85	0	0	1	0	0,45	0,5	0	0	66,7	58,86	69,73	64,00
<b>cam</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>Frank</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>devon</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>lena</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>kaii</b>	0,53	0	0,85	0	0,91 7	0	0	0	0,5	0	0	72,03	60,04	74,57	65,04
<b>Janeen</b>	0	0	0	0	0	0	0	0,45	0,5	0	0	90,5	92,86	91,36	93,75

<b>thealexque</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0	0	0	44,33	27,04	49,39	36,16
<b>barsten</b>	0	0,85	0	0	0,91 7	0	0	0	0	0	0	82,33	74,76	83,94	77,91
<b>mc</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>katia</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>kaushik95</b>	0,53	0	0,85	0,96	0	0	0	0	0,5	0	0	71,6	59,43	74,18	64,50
<b>Eddie</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>kelccccc</b>	0	0	0	0	0	0	0	0	0,5	0	0	95	92,86	95,45	93,75
<b>kellyurich</b>	0,53	0	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	1	46,13	32,04	41,94	28,04
<b>lalah</b>	0,53	0	0,85	0	0	1	0	0,45	0,5	0	0	66,7	58,86	69,73	64,00
<b>Kiana</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>kb</b>	0,53	0	0,85	0	0	1	0	0,45	0,5	0	0	66,7	58,86	69,73	64,00
<b>Carla</b>	0,53	0	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0	0	47,83	32,04	52,57	40,54
<b>Brooke</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	1	37,63	19,90	34,21	17,41
<b>Kim</b>	0,53	0	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	1	46,13	32,04	41,94	28,04
<b>Cleopantles s</b>	0	0	0,85	0,96	0	0	0,01	0	0	0	0	81,8	74,14	83,45	77,38
<b>Kinjal</b>	0,53	0	0,85	0	0	0	0	0,45	0,5	0	0	76,7	73,14	78,82	76,50
<b>Goatism</b>	0	0	0	0,96	0,91 7	0	0	0	0	0	0	81,23	73,19	82,94	76,54

<b>Kirsti</b>	0,53	0	0,85	0	0	0	0	0,45	0,5	0	0	76,7	73,14	78,82	76,50
<b>kelly</b>	0,53	0,85	0,85	0,96	0,91	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>kleet</b>	0	0	0	0,96	0	0	0	0	0	0	0	90,4	86,29	91,27	88,00
<b>Alan</b>	0,53	0,85	0,85	0,96	0,91	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>Amy</b>	0,53	0,85	0,85	0,96	0,91	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>KARLA</b>	0,53	0,85	0,85	0,96	0,91	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>Raine</b>	0,53	0,85	0,85	0,96	0,91	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>Caroline</b>	0,53	0,85	0,85	0,96	0,91	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>melissa</b>	0,53	0	0,85	0,96	0,91	1	0,01	0,45	0,5	0,17	0	46,13	32,04	51,03	40,54
<b>maria</b>	0,53	0,85	0,85	0,96	0,91	1	0,01	0,45	0,5	0	1	39,33	19,90	35,75	17,41
<b>islandlatina</b>	0,53	0,85	0,85	0,96	0	1	0,01	0,45	0,5	0	0	48,5	33,00	53,18	41,38
<b>Arindam</b>	0,53	0	0	0	0	1	0	0,45	0,5	0	0	75,2	71,00	77,45	74,63
<b>beige</b>	0,53	0	0,85	0	0,91	1	0	0,45	0,5	0	0	57,53	45,76	61,39	52,54
<b>kendall</b>	0,53	0,85	0,85	0,96	0,91	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>MCMXCII</b>	0	0	0	0	0	0	0	0	0,5	0	0	95	92,86	95,45	93,75
<b>izabella</b>	0,53	0	0,85	0	0	1	0	0,45	0,5	0	0	66,7	58,86	69,73	64,00
<b>J</b>	0,53	0	0,85	0,96	0,91	1	0,01	0,45	0,5	0,17	0	46,13	32,04	51,03	40,54
<b>ZayasS</b>	0	0	0	0	0	0	0	0	0,5	0	0	95	92,86	95,45	93,75

<b>Jake</b>	0,53	0,85	0,85	0,96	0	1	0,01	0,45	0,5	0,17	0	46,8	33,00	51,64	41,38
<b>Jam</b>	0,53	0	0,85	0,96	0	1	0,01	0,45	0,5	0	0	57	45,14	60,91	52,00
<b>Jamie</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>Janine</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0	1	39,33	19,90	35,75	17,41
<b>Jaz</b>	0,53	0	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0	0	47,83	32,04	52,57	40,54
<b>Jason</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	1	37,63	19,90	34,21	17,41
<b>SuttyBoi</b>	0	0	0	0	0	0	0	0	0	0	0	100	100,00	100,00	100,00
<b>tamara</b>	0,53	0	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	46,13	32,04	51,03	40,54
<b>debora</b>	0,53	0	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0	0	47,83	32,04	52,57	40,54
<b>javachik</b>	0,53	0	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	46,13	32,04	51,03	40,54
<b>lee</b>	0,53	0	0,85	0	0,91 7	1	0	0,45	0,5	0	0	57,53	45,76	61,39	52,54
<b>jaydah</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,5	80,71	87,73	83,13
<b>jazz</b>	0,53	0	0,85	0,96	0	1	0,01	0,45	0,5	0	0	57	45,14	60,91	52,00
<b>Jenn</b>	0,53	0,85	0,85	0	0	1	0	0,45	0,5	0,17	0	56,5	46,71	60,45	53,38
<b>Jen</b>	0,53	0	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0	0	47,83	32,04	52,57	40,54
<b>louise</b>	0,53	0	0,85	0,96	0,91 7	1	0	0,45	0,5	0	0	47,93	32,04	52,66	40,54
<b>pedro</b>	0,53	0,85	0,85	0,96	0,91 7	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>JWW</b>	0,53	0	0,85	0,96	0,91 7	0	0,01	0,45	0,5	0	0	57,83	46,33	61,66	53,04

<b>claire</b>	0,53	0	0,85	0,96	0,917	1	0,01	0,45	0,5	0,17	1	46,13	32,04	41,94	28,04
<b>Oscar</b>	0,53	0,85	0,85	0,96	0,917	1	0,01	0,45	0,5	0,17	0	37,63	19,90	43,30	29,91
<b>aubrey</b>	0,53	0	0,85	0,96	0,917	1	0	0,45	0,5	0	0	47,93	32,04	52,66	40,54
<b>enennaj</b>	0	0	0	0	0	0	0	0	0	0	0	100	100,00	100,00	100,00
<b>kyle</b>	0,53	0	0,85	0,96	0,917	1	0,01	0,45	0,5	0,17	0	46,13	32,04	51,03	40,54
<b>pato</b>	0,53	0,85	0,85	0,96	0,917	1	0,01	0,45	0,5	0	0	39,33	19,90	44,85	29,91
<b>Pastrana</b>	0,53	0	0,85	0	0	1	0	0	0,5	0	0	71,2	58,86	73,82	64,00
<b>Paulina</b>	0,53	0	0,85	0	0	1	0	0,45	0,5	0	0	66,7	58,86	69,73	64,00

Anexo B -Resultados *dataset* contas falsas

<b>AccountID</b>	<b>the account have at least 30 followers</b>	<b>has been geolocalized</b>	<b>it has been included in another's user favorites</b>	<b>it has one hash tag in at least one tweet</b>	<b>it has logged in twitter using a iph one</b>	<b>a ment ion by twitt er user</b>	<b>it has writ ten at least 50 twee ts</b>	<b>it has been inclu ded in anto her user' s list</b>	<b>2*nu mber follow ers _ numb er of friend s</b>	<b>user have at least one favo rite list</b>	<b>User verif ied</b>	<b>% fake accoun t (10 parame tros)</b>	<b>% fake accoun t (7 parame tros)</b>	<b>% fake accoun t (11 parame tros)</b>	<b>% fake accoun t (8 parame tros)</b>
<b>ffencepost</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75

<b>HunaAkamai</b>	0	0	0,85	0	0	0	0	0	0	0	0	91,50	87,86	92,27	89,38
<b>smokinlette</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>MANDO98</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>joacassi</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>clodoado</b>	0	0	0	0	0	0	0	0,45	0,5	0	0	90,50	92,86	91,36	93,75
<b>3simplysuper</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>mizzangelus</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>xlea12345</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>GabrielSouza94</b>	0	0	0	0	0	0	0	0	0	0	0	100,00	100,00	100,00	100,00
<b>Nano6620</b>	0	0	0,85	0	0	0	0	0,45	0	0,17	0	85,30	87,86	86,64	89,38
<b>miakkuching</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>AleahGreen</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>Truck_freak</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>johnyG81</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>sifluxy</b>	0	0	0,85	0	0	0	0	0	0	0	0	91,50	87,86	92,27	89,38

<b>zoedatslida</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>chenmanfu</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>anders200</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>darkhartz</b>	0	0	0,85	0	0	0	0	0	0,5	0,17	0	84,80	80,71	86,18	83,13
<b>cleo_junior</b>	0	0	0,85	0	0	0	0	0	0	0	0	91,50	87,86	92,27	89,38
<b>cleber_1496</b>	0	0	0	0	0	0	0	0,45	0,5	0	0	90,50	92,86	91,36	93,75
<b>sketayo</b>	0	0	0,85	0	0	0	0	0,45	0	0	0	87,00	87,86	88,18	89,38
<b>westdean60</b>	0	0	0,85	0	0	0	0	0	0	0	0	91,50	87,86	92,27	89,38
<b>adie_austria</b>	0	0	0,85	0	0	0	0,01	0	0	0	0	91,40	87,86	92,18	89,38
<b>dieguinhocn</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>carmil9698</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>HabboJoBros</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>ee0012</b>	0	0	0	0	0	0	0	0,45	0,5	0	0	90,50	92,86	91,36	93,75
<b>chussin_20</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>EurwynRoyce</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13

<b>Ernesto2277</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>Hotdimo</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>jean_christien</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>belamri12312</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>vampiro_666819</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>epo58</b>	0	0	0	0	0	0	0	0,45	0,5	0	0	90,50	92,86	91,36	93,75
<b>luay10</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>vega_alma7</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>kamel009</b>	0	0	0	0,96	0	0	0	0,45	0	0	0	85,90	86,29	87,18	88,00
<b>Thatag15</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>IkarosLove</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>rul085</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>paynevscastle</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>EightSevent</b>	0	0	0	0,96	0	0	0,01	0	0,5	0	0	85,30	79,14	86,64	81,75
<b>donsilabas</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>Lamarcio</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75



<b>verionboss</b>	0	0	0,85	0	0	0	0	0	0,5	0,17	0	84,80	80,71	86,18	83,13
<b>Naty_Gazzano</b>	0	0	0	0	0	0	0	0	0	0	0	100,00	100,00	100,00	100,00
<b>kittykat229</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>Monzter12</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>ady22hh</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>yunfeng7456203</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>Rafitaone</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>H8M3R8</b>	0	0	0	0	0	0	0	0,45	0,5	0	0	90,50	92,86	91,36	93,75
<b>angelique19999</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>Dans_gokiel</b>	0	0,85	0	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>rodrigootrancee</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>0412madspell</b>	0	0	0,85	0	0	0	0	0	0	0	0	91,50	87,86	92,27	89,38
<b>fonk22</b>	0	0	0,85	0	0	0	0	0	0	0	0	91,50	87,86	92,27	89,38
<b>KliinOo</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>Clau_Ulloa</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13

<b>timeoutmiranda</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>DustinWats on3</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>lailadayanti</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>rockdrummer1990</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>Normandk</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>hackedbiz</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>lele5m</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>saul460</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>tungmaye</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>lsmith86</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>ntimonesjr</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>3rickm0nt</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>nasantillan</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>avel01</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>mardonimatos</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13

<b>ZeGuilhermo</b>	0	0	0	0	0	0	0,01	0,45	0,5	0	0	90,40	92,86	91,27	93,75
<b>xiiLOVEY OUU_</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>realLeDuc</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>ninenutkub</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>tu_baby_star</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>rasyefare</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>f3rn4ndo1</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>F_FreSshMusic</b>	0	0	0	0,96	0	0	0	0,45	0	0	0	85,90	86,29	87,18	88,00
<b>Andres19781157</b>	0	0	0	0	0	0	0	0	0	0	0	100,00	100,00	100,00	100,00
<b>55_OeMeR_55</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>benji_sib</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>nouiouaha</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>C4rL1T0sSS</b>	0	0	0,85	0	0	0	0	0,45	0,5	0	0	82,00	80,71	83,64	83,13
<b>ajie_ramadhan</b>	0	0	0,85	0	0	0	0	0	0,5	0,17	0	84,80	80,71	86,18	83,13
<b>joaozikamuleque</b>	0	0,85	0,85	0	0	0	0	0	0,5	0	0	78,00	68,57	80,00	72,50

<b>Xuaner_96</b>	0	0	0	0	0	0	0	0	0,5	0	0	95,00	92,86	95,45	93,75
<b>agna_7</b>	0	0,85	0,85	0	0	0	0	0	0,5	0	0	78,00	68,57	80,00	72,50
<b>argo_br</b>	0	0	0	0,96	0	0	0,01	0	0	0,17	0	88,60	86,29	89,64	88,00
<b>MMooDeW W</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>krolbranco 1</b>	0	0	0,85	0	0	0	0	0,45	0,5	0	0	82,00	80,71	83,64	83,13
<b>ricuarachel i</b>	0	0	0,85	0,96	0	0	0	0	0,5	0	0	76,90	67,00	79,00	71,13
<b>gui488</b>	0	0	0,85	0	0	0	0	0	0,5	0	0	86,50	80,71	87,73	83,13
<b>joshy2314</b>	0	0	0,85	0	0	0	0	0,45	0,5	0	0	82,00	80,71	83,64	83,13