**Impact of bots in our life**

First bots were created for good purposes, like good tools that will help online businesses to keep their content on track, search engines, etc. We can mention here Google bot, Amazon bot, Twitter bot, Facebook bot, etc. With the evolution of online business, bots are used also for malicious purposes. Recently with the pandemic situation, most of the businesses are operating online and this has its owns advantages and disadvantages. We can buy all essential things online and this is good for us, but on the other side, the business itself is facing a threat. Their data is vital and there are several bots who would like to steal it.

It doesn’t really matter if you are a CEO of a hundred companies, a data analyst, a marketer, a journalist, we all drive our strategies and insights from data. Data is the key that will make the difference. There are different bots that visit your website. Web scrapers, Account Takeover/Credential Stuffing bots, Sneaker bots, ticket scalping bots, etc. We will focus on Web Scraping and ATO/Credential Stuffing.

**What is Web Scraping?**

Most people have done once in their life copy and paste manually, and this is the easiest way of scraping.

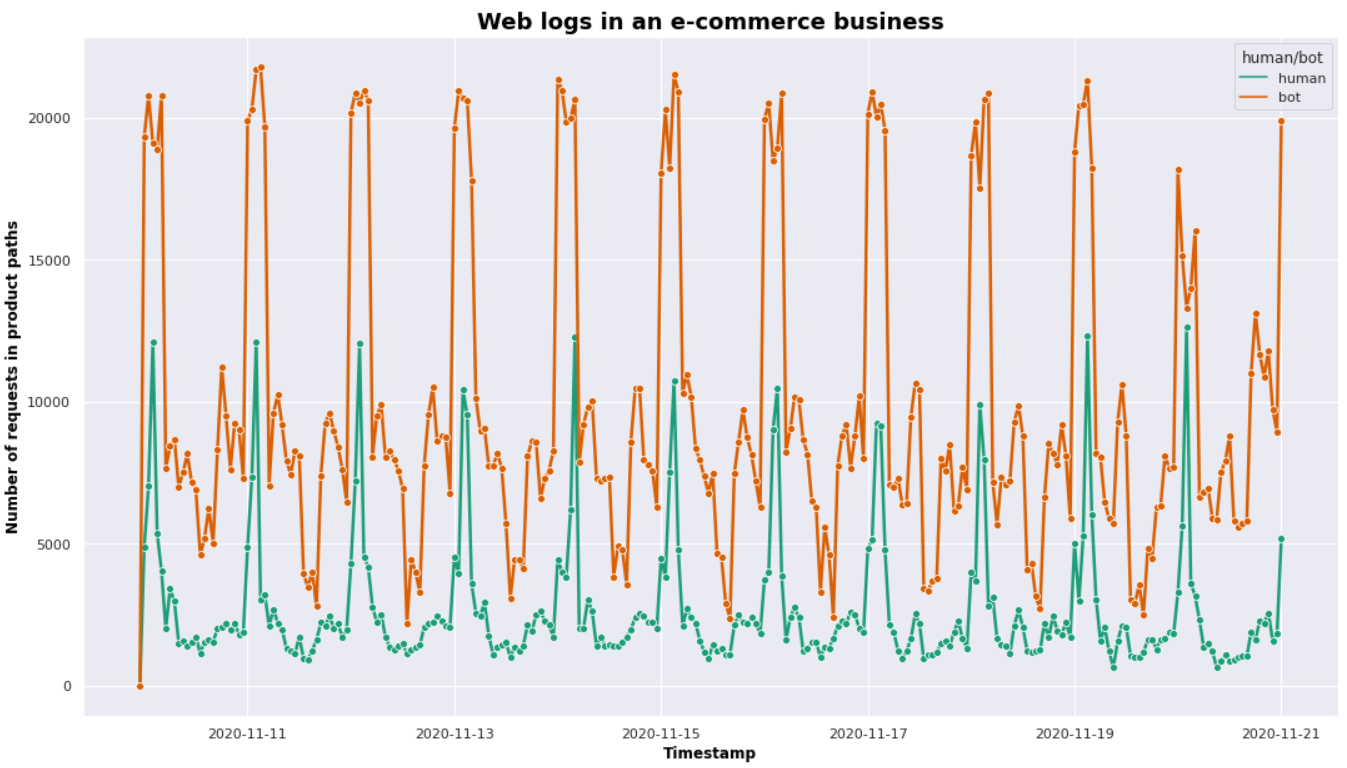
With the development of technology and online business there are hundreds of pages that you can’t just copy and paste. Today there are a lot of tools, library’s that can help to do web scraping. We can mention here some python libraries: **Requests (HTTP for Humans), Ixml, Beautiful Soup, Selenium, Scrapy, etc.**

[Web scraping](https://prowebscraper.com/blog/the-ultimate-guide-to-web-scraping-for-non-programmers/) is data scraping a process of automating the extraction of data in an efficient and fast way. With the help of web scraping, you can extract data from any website, no matter how large the data is, on your computer. After gathering this data, you can parse this data and create useful insights for different purposes. [More information](https://www.netacea.com/)

Your website can be scraped for many reasons: your competitor can scrape your website to clone your products or to do price comparing, different educative institutions can scraper your data for analytic projects.

**Example of product scraping in an e-commerce business**

Usually, 40% of the traffic on your website is coming from bots. In this example, we have presented a time series that shows the number of requests in product paths coming from humans vs bots.



As we can see from the graph above the number of requests coming from bots is very high compared to humans. In this case, 79% of the traffic in the product paths is coming from scrapers and only 21% from humans.

**What is Account Takeover/Credential Stuffing?**

Account Takeover/Credential stuffing is a threat for everyone that have an online account.

Credential stuffing is a type of cyberattack where stolen account credentials typically consisting of lists of usernames and/or email addresses and the corresponding passwords are used to gain unauthorized access to user accounts through large-scale automated login requests directed against a web application. [More information](https://www.netacea.com/)

While account takeover fraud is using bots to access a legitimate user’s login details (username & password) to take over their online account. They can get access to the victim’s bank, financial site, e-commerce site, or other type of accounts. They can do credit card theft or spend an unauthorised monetary for shopping. Some of more used credential stuffing tools are: **SentryMBA, Vertex, STORM, Snipr, Account Reaper, Hydra, Cr3dOv3r, PrimeKiller, Joyn, etc.** Credentials are usually sold on the clear web, on many different online forums and market places.

**Example of ATO/Credential Stuffing attacks in an e-commerce business**

In this plot, we have presented a time series that shows the number of requests in login paths coming from human’s vs bots. Most of the spikes we see in the login are credential stuffing attacks, but we can also see other spikes which can be part of a testing and is not required to be mitigated. There are some kind of attacks: sophisticated or unsophisticated. A sophisticated attack is in terms of distribution by IPs, data centres, countries, etc, while unsophisticated attacks may be very massive in terms of login attempts but very less distributed. Another challenge of attacks is the volume of requests.



In this case, approximately 64% of the login activity is coming from bots that are trying to gain access to some accounts and use it for malicious purposes. In order to prevent their customers from this threat, companies are investing in advanced bot management.

**How can Machine Learning help to identify bots?**

In this case, we are in a classification problem where we have to predict if the upcoming user is a human or bot. To do that we can use supervised, unsupervised learning, or anomaly detection. In supervised learning, the algorithm will learn based on the previously labelled dataset, providing an answer key that the algorithm can use to evaluate its accuracy on training data. In an unsupervised model, the algorithm tries to make sense of unlabelled data by extracting features and patterns on its own. In anomaly detection, we can do some data analysis and identify rare actors that might be suspicious, and also change the form of the data. In supervised learning, we can mention these algorithms logistic regression, decision tree, neural network, random forest, gradient boosting, support vector machines, etc. In unsupervised learning, there are algorithms like K-means for clustering problems, Principal Component Analysis, Independent Component Analysis, etc. If we have successfully predicted the user then we will be able to mitigate and probably block it. To detect bots in the web scraping case above we have used supervised learning and in the account takeover/credential stuffing we have used unsupervised learning and anomaly detection.