
# Machine Learning
# (32513)

# Anti-Bot: Improving the Fairness of Limited-Release Sneakers

**Yilanqing Zhong**

**12828277**

# Table of Content

# Introduction

The global sneaker market cannot be underestimated. According to Grand View Research(2018), the market is expected to reach $95.14 billion by 2025. Nowadays, due to the increasing participation of the young population in sports activities and increasing consumer demand for limited sneakers, the secondary market for limited-release sneakers is estimated to be worth more than $1 billion (Wolff & Rega 2016). Therefore, driven by interests, a large number of BOT software appear. Although there are many basic ways to stop bots, they cannot stop bots effectively because bots evolve over time.

# Aim, objectives and expected results

## Aim

The primary aim of this project is to make the sale of limited-release sneakers fairer, by combining reCAPTCHA with face recognition and abnormal order cancellation system.

## Objectives

The overall aim is broken into three objectives, as shown below:

- Purchase the most popular or best BOT software in the market as samples to understand how bots works and how bots successfully avoid the existing anti-bot approach.
- Set up face recognition
- Collect and analyze purchase records to distinguish humans from bots, and build training model.
- Validate and deploy the anti-bot approaches.

# Background

Some years ago, when consumers used a computer to buy sneakers, they would be tested. For humans, this should be an easy test; for a computer or BOT, the test should be almost impossible to solve. This kind of test is known as the CAPTCHA, which

stands for the Completely Automated Public Turing Test to Tell Computers and Humans Apart (Strickland 2008). Captcha tests could be found on many shopping websites. The most common form is an image of several distorted letters with chaotic and textured backgrounds, requiring the user to enter the correct letter in the box. If the input letter matches the letter in the image on the screen, the test is passed. CAPTCHA tests help determine which users are real people and which are bots. However, with the development of algorithms, the traditional CAPTCHA can be solved by computers gradually. According to Mori & Malik (2002), distorted images and added chaotic textured backgrounds could be already be correctly identified with high probability. Moreover, as the complexity of images increased, it became increasingly difficult for humans to recognize images.

After reCAPTCHA was acquired by Google in 2009, Google started to phase out captcha identification in 2014. In 2017, Google announced that it was getting rid of CAPTCHA. The latest version, the "Invisible reCAPTCHA v3", removes the box-ticking. It does not interrupt users, and it monitors how users have interacted with a website, such as how they clicked the "Checkout" or "Place Order" button, how they have moved the mouse around and analyzed browsing habits and data to determine whether the testee is human or bots (Google Webmasters 2018). If the testee seems suspicious, the testee will see a grid of nine photos, requiring the testee to pick the corresponding images.

However, BOT users still have a lot of ways to solve the reCAPTCHA problem. And this is the reason why the other two approaches are proposed to combine with reCAPTCHA.

As for face recognition, it has been widely using these years. Face recognition technology is a kind of biometric recognition technology which is more accurate than fingerprinting. Firstly, the focus of data collection is confirmed through face positioning, and the image capture system takes photos to process the key data which

is extracted by the system. Then the information in the database is compared by the search system, so as to complete face recognition and distinguish the identity of the identified. However, it is worth mentioning that most of the current facial recognition technologies are based on a two-dimensional plane and rely on dozens of feature points on the image to generate a set of feature values. Then, each recognition will be compared with the first input of feature values, and a certain accuracy will be considered as a successful match. This approach has a better guarantee in the speed of identification, and accuracy is not bad, but the security can not stand the test. However, Apple's Face ID uses image recognition technology based on three-dimensional imaging, which is more secure than conventional two-dimensional image recognition. In this project, face recognition will first be considered for mobile applications due to the hardware requirements. Therefore, face recognition can be set up with reCAPTCHA as a new anti-bot solution.

Nowadays, many shopping sites, such as Nike and Supreme, place a lot of different shields against bots before customers place orders, such as reCAPTCHA. But these approaches cannot totally solve the problem that there are still many bots that can checkout at a speed that humans cannot. As a result, face recognition and an abnormal order cancellation system are proposed in this project. The abnormal order cancellation system is a method combined with machine learning to check orders and cancel abnormal orders.

# Research Project

## Project significance

What takes human dozens of seconds to complete the order (this does not consider network congestion into account), the BOT can do it in merely one second. Furthermore, by deploying hundreds of proxies, BOT users are able to purchase hundreds of sneakers at the same time (Steinberg 2018). In the long run, BOT users will monopolize the limited amount of sneakers and have a negative impact on brands

and dealers, such as the proliferation of fake products, the loss of consumers' confidence. Therefore, anti-bot is imperative, and multi anti-bot approaches are necessary.

## Project innovation

Currently, most shopping websites use Shopify as the website as the framework, and reCAPTCHA is the primary method of anti-bot. Although face recognition is widely used in many areas, it is not used as an anti-bot approach. Moreover, with the advancement in technology, some bots now have a human mode in addition to normal mode; that is, bots can imitate human behavior to enter information, click and move the mouse to avoid the detection of anti-bot program. Also, some genuine users are likely to be detected as bots by mistake, thus being penalized and restricted. At the same time, very few websites implement the accurate abnormal order cancellation system. Many genuine users' orders may be somehow canceled. Therefore, the simultaneous implementation of effective multiple anti-bot approaches is also an innovation of this project.

## Project outline

In order to effectively block Bots, we need to be familiar with Bots; that is, understanding how Bots work. Therefore, it is a good idea to directly purchase the most popular or best bots on the market as samples. Software engineers can simulate the purchase process by operating the bots to understand the mechanics of Bots and find out the current vulnerability, namely how the bots manage to evade reCAPTCHA's detection. Software engineers can set up a face-recognition program on a random page and determine that testee is the user by taking a photo of the user in advance and comparing it to a real-time photo. This requires that each user has an account and an uploaded photo.

Limited-release sneakers by large websites such as Nike and Yeezysupply will generate thousands of purchase records. It is undoubtedly time-consuming and

inefficient to check abnormal orders manually. Therefore, this process can be done with machine learning. In detail, the training of order data through the neural network can assign weights to different attributes, such as order completion time, payment information and whether the delivery address is repeated; then, the threshold can be determined to identify genuine users and bots. Finally, evaluate the performance of the algorithm and deploy it for practical use.

It is important to note that the data for each release may differ. Although training data can be collected through internal testing, but the reality is often more complicated because when the sneakers are sold in limited quantities, the network tends to be jammed due to the influx of thousands of users, which will affect the order completion time.

### Project timeline

- Research on bots samples: 2 months
- Collect data and train model: 6 months
- APP development: 3 months
- Deployment and validation: 1 month

### Expected results

- Create a test that humans can easily pass, but BOT cannot, i.e., face recognition.
- Even bots help pass the test quickly, limit the number of purchases.
- The misidentification rate of human and BOT behavior is less than 1 percent.

In the long run, the unfairness will lead to the proliferation of fakes and lead people not to buy genuine products. Therefore, brands like Nike and Adidas will benefit from improving fairness in limited-release.

## Budget

| Category | Cost($A) |
|---|---|

| **Salary (annual)** | |
|---|---|
| **Project Manager (1)** | **120,000** |
| **Software Developer (2)** | **220,000** |
| **Data Analyst (2)** | **250,000** |
| **Assistant (3)** | **150,000** |
| **Network Engineer (2)** | **180,000** |
| **Infrastructure** | |
| **Database** | **5,000** |
| **Server (2)** | **10,000** |
| **Computer (12)** | **30,000** |
| **Software (including bots samples)** | **10,000** |
| | |
| **Overall Cost(Salary & Infrastructure)** | **975,000** |
| **Contingencies (10%)** | **97,500** |
| **Total** | **1,072,500** |

*Table 1: Project budget*

## Personnel

**Project Manager:** supervise and manage this project.

**Software Developer:** Research on bots, App development, deploy face recognition, and combine with reCAPTCHA, deploy the abnormal order cancellation system.

**Data Analyst:** develop the algorithm for abnormal orders cancellation system.

**Assistant:** Assist software developer, data analyst, and network engineer.

**Network Engineer:** build servers and maintain networks.

## Reference

Google Webmasters 2018, *Introducing reCAPTCHA v3,* Youtube, viewed 8 October 2019, <https://www.youtube.com/watch?time_continue=138&v=tbvxFW4UJdU>.

Grand View Research, Athletic Footwear Market Worth $95.14 Billion By 2025 | CAGR: 5.1%, viewed 8 October, <https://www.grandviewresearch.com/press-release/global-athletic-footwear-market>.

Mori, G. & Malik, J. 2003, 'Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA', *Institute of Electrical and Electronics Engineers,* vol. 1, pp. 134-144.

Wolff, J. & Rega, S. 2016, *The sneaker resale market is worth a massive $1 billion and these are the sneakerheads driving it*, Business Insider Australia, viewed 8 October 2019, <https://www.businessinsider.com.au/inside-the-billion-dollar-sneakerhead-resale-economy-2016-5?r=US&IR=T>.

Steinberg, L. 2018, *The Profitable Hidden Sneaker Marker*, Forbes, viewed 8 October 2019, <https://www.forbes.com/sites/leighsteinberg/2018/09/17/the-profitable-hidden-sneaker-market/#4d78563f5925>.

Strickland, J. 2008, *How CAPTCHA Works*, howstuffworks, viewed 8 October 2019, <https://computer.howstuffworks.com/captcha.htm>.

## Video pitch

https://www.youtube.com/channel/UCL_Hvp8ns9JUkOAbzQ26enQ