Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Master Degree in ICT for Internet and Multimedia

FittingLabs: study case, data analysis and development of an e-commerce site through machine learning

Supervisor:
Prof.ssa Federica Battisti

Candidate:
Alessio Lovato
2021115

Academic Year 2021-2022

"There's a time and place for everything, but ~~not~~ now."

Prof. Oak

Abstract

Ever since the Internet has become a tool of everyday use, the development of online shopping platforms became a necessary implementation for all major companies. The convenience of buying anything anywhere is a habit that users have appreciated more and more. However, during the pandemic of recent years, it became a necessity.

The online sale of clothing has changed over the years, but the approach that the user takes during the purchase phase has remained unchanged. Nowadays, the majority of the companies are offering a free return when the purchase did not meet the user expectations. The intention of this research is to explore and develop a new approach to online shopping, providing shoppers with a new way to select the clothing item that best suits their needs. An innovative purpose such this would increase the media impact, making an online purchase more reliable and accurate, limiting the number of returns made by buyers.

The purposed approach started with an analysis of a data collected from a wide set of users. Those were elaborated and analysed to find out if it was possible to describe some correlations. Different methods were then explored to solve the problem. It was used a CF approach, a CBF approach and finally a Regressor to predict the outcome that could be suggested to the user.

This research highlights how already in an initial phase this approach seems to be optimistic and how it is possible to exploit machine learning to provide a more reliable service for users, and advantageous for companies.

Contents

List of Figures

Chapter 1

Introduction

Since the beginning of the 19^{th} century, mountain climbing was practiced as a recreational activity. Italy, France and England were the first locations where climbers start to passionate about climbing. Alps were indeed one of the favorite destination for all early mountaineers. In particular Dolomite were one of the birthplace of rock climbing.

In the beginning, climbers were driven by the need to reach unexplored rocky areas and the mountain top. However, as time went on, this need became less and less, as the unexplored areas became less and less. The desire to climb remained, however, very vivid, such that climbers began to join in the sports community to share this passion together.
There were no sophisticated or specific equipment at the beginning. Climbing bolts were first used in the middle of 1900, while only later in the 1960s in USA climbers pulled gears to accomplish aid climbing tactics[1].

Early rock climbers were used to wear common heavy-soled mountaineering boots with metal cleats. Only in 1930 an Italian entrepreneur invented the first rubber sole, called Vibram. This particular rubber sole allows exceptional grip and safety with respect to the boots.

Since then, Italy has been considered a leader in technical climbing clothing. In particular, the two most famous brands are Italians, leaders as brands and tech used while creating new climbing shoes.
However beginners and climbers in general encounter numerous difficulties when purchasing shoes. The necessarily narrow and enveloping structure of the climbing shoes requires it to be particularly similar to the anatomical shape of the foot. Buyer must reach a

[1]**Aid climbing tactics:** it is a style of climbing in which standing on or pulling oneself up via devices attached to fixed or placed protection is used to make upward progress.

physical store and try multiple types of shoes and different sizes before to be able to buy the right one. Even this approach won't eliminate all the issues. The construction material of each individual model influences how much the shoe is able to widen at first, and how much it deforms after months. This variables should be taken into consideration by the climber while buying a new set of climbing shoes, including the experience of the seller, who may or may not suggest a specific model size. Finally users need to accept that not all models would be available in local retails, some brands are not always covered by the local store.

As a climbing shoe e-commerce it's mandatory to implement a refund and return policy to meet the buyers need, who rarely are able to go for the right size immediately. Sometimes a user interested in buying a pair of shoes online makes a wrong purchase due to inaccurate or misinterpretable guidelines available on the site.

Where did the need to develop a new way of doing e-commerce come from? Before coming to this conclusion, several questions were asked and many answers given.

The beginning of this journey started analysing and investigating if this problem could be solved and lead to a new concept of online shopping. Is it possible to build something helpful for expert climbers and new users? Is it possible to provide a reliable plugin able to suggest what to buy and which size?

Chapter 2

Virtual Fit

The exploration of the necessities and the issues that people need to face up everyday are very hard. As users and buyers, humans adapt and fit with the environment they deal with. The investigation process of that mechanism is indeed very difficult. It could take more than a while. Also the definition of doable is not always straightforward.
Shoe e-commerce more than the others lack on buyer care, providing an unreliable solution to choose the right shoe size.

Analysing the major wearable e-commerce currently available online it is possible to realize that the physicality of a user is marginally taken into consideration. Fitting guidelines are very general. The physical features of every single human being are instead relevant once it has to deal with wearing.
The human body influence significantly the choice dealing with the next purchase, and major companies are not trying to adjust them self based on buyers. An even more delicate topic on this side is correlated with the online shopping of shoes. Generally shoes are very specific due to the extensibility of the material or the shape of the foot correlated with the shoe stitching.

Virtual Fit was born as a solution to introduce a new approach to this issue. It is first of all necessary to understand the environment in which you are venturing and what are the possible critical points to be solved.

2.1 Study case

At the beginning it was analysed the biggest company currently leader of the shoes market. It was observed Nike e-commerce situation over the last 3 years. The increasing number

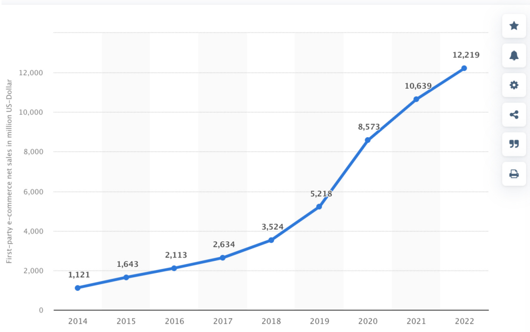

Figure 2.1: E-commerce net sales of nike.com between 2014 and 2022 [6].

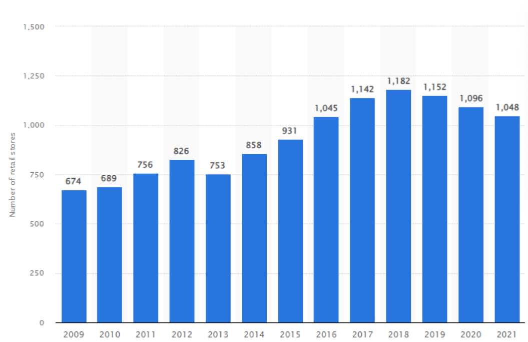

Figure 2.2: Total Nike retail stores worldwide from 2009 to 2021 [7].

of online purchases are significantly showing that the user are faster moving toward a new concept of shopping. The number of online purchases duplicated over the past 4 years. Together with this significant information it could be found that the number of retails are diminishing after a huge increasing between the 2014 and the 2018. Also Nike introduce a variety of new products with discounts while buying through the e-commerce. It is also promoting new exclusive products that will only be available through online purchases.

The significance of those outcomes are though very relevant. Body Labs company has modeled that there is a 62 billion worth of apparel and footwear every year. The 57% of the footwear is generally returned, since the size won't fit the buyer. On the other side, the 51% of the potential sales are based on fit confidence [2]. It is relevant since that generally the return option is fully in charge of the company. Though it's not considering that dealing with returning a pair of shoes is a hassle for the buyer.

To underline how this economy is hardly sustainable, recently some companies are moving forward adding a symbolic payment while returning a wearable product. They decided to add a small fee to positively accomplish a return within the established deadline [11].
The main reason is clearly of an economic nature. The recent boom in online shopping, and the consequent decrease in sales in stores, has made this change necessary. Users buy

the same product in multiple sizes. They prove it, and return those who do not convince them.
If this is not enough, the environmental impact of a behavior like this can also be considered.

2.2 Virtual Fit

Virtual Fit aspires to be the next generation online shopping experience. As said users need to be very flexible while dealing with online shopping. Now a days, once approached in an online e-commerce, the buyer need to investigate and understand how to fit the information provided by the website to their needs. Generally it's necessary to measure the length of the feet or to print guidelines in a paper to have a general understanding if the product would probably fit well my feet.
Virtual Fit aims to cancel this process. Each single user would be followed through a step-by-step procedure who will avoid this interpretation procedure. The application will receive as input some images from the users. Given that, the algorithm would be able to elaborate the image into a 3D representation and to provide the best fitting size to the buyer.

The Virtual Fit solution is intended to be seen as a 4 steps technology: identify the item I intend to buy [Figure 2.3], find the model in an e-commerce [Figure 2.4], it will be used VirtualFit to choose for my right size [Figure 2.5] and finally buy the item with the awareness that it would be the right one [Figure 2.6].

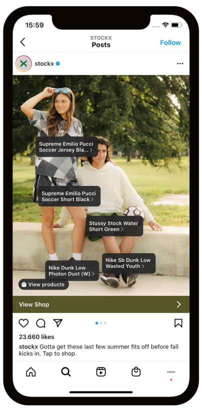

Figure 2.3: Step 1.

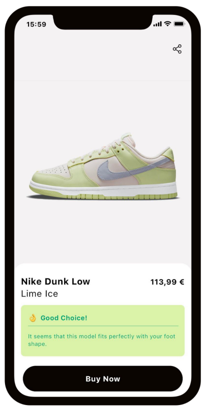

Figure 2.4: Step 2.

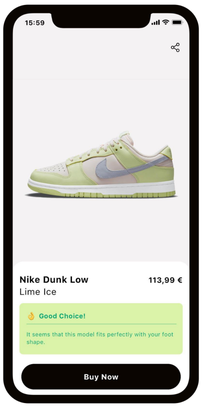

Figure 2.5: Step 3.

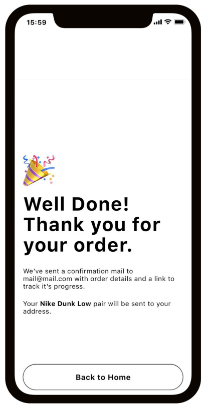

Figure 2.6: Step 4.

2.3 Technologies involved

Dealing with a piece of technology it is important to take into consideration how to approach the problem. So, at the beginning, it was planned a possible solution to outcome with the right size of the shoe that the user intends to buy. Dealing with an anatomical body part analysis, the first approach that came into mind resemble the 3D feet reconstruction. This approach highlight the possibility to retrieve the feet measures and to adjust them based on the shoe that the user is going to buy.

The first step would retrieve a free hand video or a sequence of images. Here it is considered that the image/video taken has a good quality. It should be necessary to guide with an UI the user: this would improve the UX during the process of acquiring the image/video. It will also increase the probability to receive a consistent set of data that can be used.
Given that it's necessary to extract the feet interesting points. It could be done extracting point cloud data from the image and generate the 3D meshes. Point cloud could be easily retrieved through the LiDAR[2], currently available in Apple products. On the other side it could be possible through a direct photogrammetry algorithm where cameras should capture multiple images from different angles. It is possible to reconstruct the space as a 3 dimensional later on.
To isolate the foot from the actual environment, it would be possible to assume that it is the biggest closed surface on scene.

[2]**LiDAR:** it is a technology able to retrieve the space location of each point seen on the surface of an object. It can target an object/surface and it can measure the time spent by the light reflected from the object to return to the receiver.

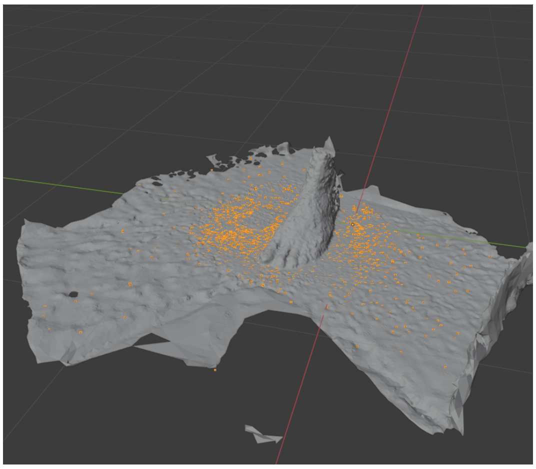

Figure 2.7: Reconstructed mesh with floor detection and rectification.

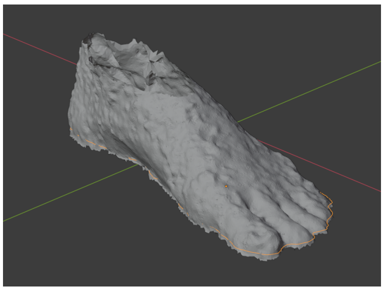

Figure 2.8: Single foot extraction.

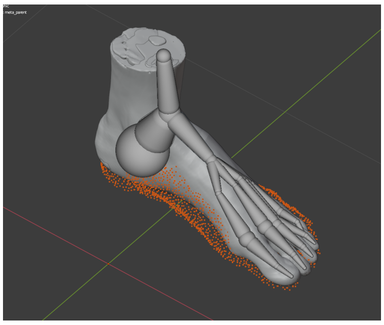

Figure 2.9: Anatomical regression.

At this stage it could be possible to use an Anatomical regression[3] on the model of a foot. The retrieved data from the user could be adjusted to the model into a representation of a feet.

Since it would deal with regression, it would be necessary to own an anthropometric set, that should be able to extract the most probable distribution of tissues compatible with the 3D mesh retrieved.

To increase the number of constraints of the tissues it maybe should be necessary to have multiple images taken into different positions, to understand the available bounds. Finally it would be necessary to validate that the outcome of the regressive model would be correct. This should be done manually from experts capable to deeply understand the outcome. To validate it could be ask for help an orthopedic.

At this point there would be 3 different relevant outcomes:

1. Anatomical measurements (metatarsal joints, ball arc, etc.)

2. Shoe last to Foot registration and distance map. This would allow to understand the

[3]**Regression:** it is a statistical technique that relates a dependent variable to one or more independent variables.

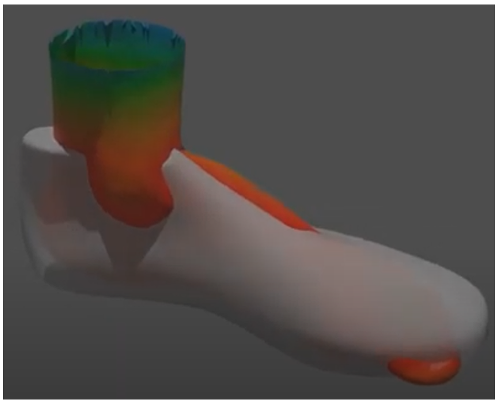

Figure 2.10: Shoe last to Foot distance mapping.

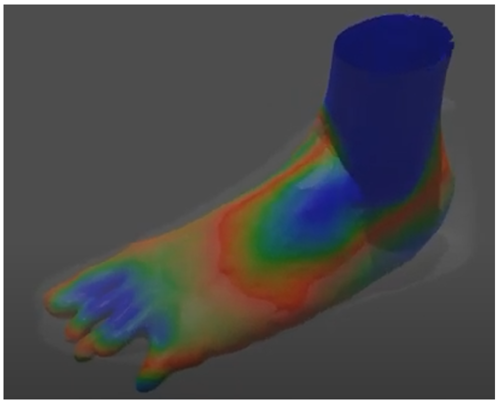

Figure 2.11: Shoe last to Foot distance mapping.

best size of the Shoe last would fit the retrieved foot [Figure 2.10][Figure 2.11].

3. Shoe last to Foot physically based registration and stress analysis: given a reasonable physical model of the last of the shoe and a reasonable physical model of the foot it is possible to logically physically deforms the representations to fit them together.

To compute the comparison between the foot and the shoe, it must be available the parametric model of the shoe together with the data sources. So it should be necessary to have those information for all the desirable models. This would scale based on the interest of a specific product.

2.4 Impediments

From this building process it was possible to identify some difficult elements: the impediments and the external data sources.
The impediments study highlighted how hard could be to retrieve an anthropometric set of feet, the process used to validate the regression and estimation of the accuracy, the extraction of a reasonable model of the shoe's last and foot and output the foot-to-shoe contact points. Speaking about the external data it's both important to recover at the best 10 to 100 pictures sequence or a video. The images or the video should be taken along 360°. To retrieve the parameters of a shoe model, it would be necessary to have the model itself. It would be necessary to analyse it, extract the structure to be later used during the fitting process.

Once it was possible to identify all the weak points of the process, it was clear and overview of the work that should be done.

Chapter 3

FittingLabs

Virtual Fit concerned about sneaker, trainers and everyday shoes. Virtual Fit's study case found a company interests, who saw the possibility of adapting this process specifically to the field of climbing. This would lead to the introduction to some other specifications and process to take into consideration. As said in the introduction, technical wears should be extremely precise and adapt very tightly and precisely to the user needs. Sport wearing in particular would introduce new layers to the previous analysis. It is necessary to take into consideration not only the comfort but also the performances.

FittingLabs is the consequence of Virtual Fit, and it would focus on the study of climbing shoes. It takes straightforward the considerations done for Virtual Fit. Nevertheless it is necessary to gather new information and features about the world of the climbing shoes.

3.1 Climbing shoes structure

While joining the world of climbing shoes, it is generally not taken into account that climbers are using extremely specific and advanced technical clothing. This sport sees the climber physically struggle against the gravity alone: the only factors that come into play to help the sportsman are its fingers, its feet toes and arms. Although this is partially true, over the years technology has developed boots capable of adapting to the specific situations a climber may find himself in. Not only that, the boot can be specifically developed for specific routes, walls and abilities of the climber.

3.1.1 Shoe structure

Climbing shoes could be divided into 16 different areas:

Figure 3.1: Anatomical structure of the upper side of a climbing shoe [1].

1. The **heel loops** are two elastic laces at the back of the shoe that would help climbers to slip their feet in.

2. Climbing shoes are supposed to be worn barefoot, a **lining** is added to increase the comfort.

3. The **tongue** and the **tongue tab** are used to open more the opening and to help climbers to slip the foot in the shoe. This is available only in slippers and velcro closure shoes.

5. The **closure system** could be of three different types: laces, velcro or slipper.

6. The **upper** is just made from different material: it could be both true leather both synthetic leather. The leather could be both lined or unlined.
   Different materials have their pros and cons to be taken into consideration before buying a new model.

7. High performance shoes are built with a **toe rubber** on the top of the toe. It allows more strength to the climber while hooking.

8. Closer to the toe rubber the **rand** is covering all the toes with a different material with respect to the Upper: the rubber.

9. The **heel** coincides with the heel itself and it's made of rubber too. It allows more grip to climbers when it is needed.

10. The **Edge** of a climbing shoe is an important part that coincides with the most advanced part of the shoe and which allows the climber to lean on the smallest ledges in the rock.

Figure 3.2: Anatomical structure of the lateral side of a climbing shoe [1].

11. A huge impact on the choice of the right shoe is up to the **last**. It specifies how the shoes had been built. Both the shape and the width of the shoe itself is determined by the last.

12. The **shape** is a consequence of the shape and represents how curved is built the shoe. Shoes could be built with a flat-sole or with an aggressive downturned sole. In the first case you have a comfy approach while in the second one the climber is looking for the performance approach, the strength during the climb.

13. Since climbers need to feel tightly wrapped from the shoe, the **heel cup** helps to tightly immobilize the heel.

Figure 3.3: Anatomical structure of the lower side of a climbing shoe [1].

14. Heel could be used to execute an heel hook[4]. The material on the back side of the shoe, the **heel rubber**, is intended to allow climbers to feel comfortable while executing this action.

15. The **midsole** is a rubber material covering the arch area. If it is not present the shoe would be more comfortable. On the other side a rubber midsole would increase the overall grip and performances of the model.

16. The **sole** refers to the area underneath the toes and the knuckles. It is made of rubber to provide adherence to the surface.

The user once decided to buy a new model has to cross all those information with the usage intended. The characteristics here listed can't be divided as good or bad, they would be helpful and useful in specific situations, while won't in others.

3.1.2 Shoe shape

Another indispensable feature that should be taken into consideration was previously describe, that is the shape of the shoe. Climbing shoes are made to wrap and fit very tight the feet but it is simultaneously a constraint for the climber. The climbing shoe has to bind the foot in the right position, to allow the best posture to unload the strength on the wall. The posture of the feet are partially determined by the shape of the shoe.
Climbing shoes comes mainly into 3 different shapes:

Figure 3.4: Neutral.

Figure 3.5: Moderate downturning.

Figure 3.6: Aggressive downturning.

Neutral shoe [Figure 3.3] owns this comfortable flat shape together with a stiffer rubber that won't allow the climber to lay on a small overhang. Nevertheless it is possible to flex the sole directly on the rock to use a wide area of the shoe to have friction. This shoes generally comes with a straight asymmetry since it won't require the climber to lay on the big

[4]**Heel hook:** it is a maneuver that elevates your heel above your torso, allowing you to propel your body weight with your leg muscles.

toe. Those shoes are generally used for smearing[5].
The slightly downturned shoes [Figure 3.4] are shoes that are overall good and flexible. The sole is thinner with respect to the previous one. It allows to climb different types of routes with different techniques: slabs[6], cracks[7], mild overhangs[8] and multi-pitch[9].
The aggressive type of climbing shoes are the shoes that would return the best performances on a short term climb. They are painful since the curvature is emphasized, the sole is thin to allow climbers the best feeling while climbing. It allows to lie on a very small overhangs of rock. The rubber is extremely sticky and returns a better sense of security. On the other side those shoes are not good for a long-term climb since they are painful and create a lot of tension on the heel and on the toes. Once a new user is intended to buy a new pair of climbing shoes has to take into consideration all those information.

3.1.3 Climbing routes

Climbing shoes are indeed made to fit better in over a specific wall. There are different routes that a climber could climb and each of them require climbing shoes with different features.
The three main routes are: boulder, crag and multipitch. The boulder route is intended a style of climbing which is not using ropes and harnesses. In this case the climbing would be short, a few meters, but simultaneously very hard and demanding. The climber needs to use a safety mats just below the path in case he/she falls down.
The crag route is intended as a generally vertical, smooth rock with cracks following the vertical direction. The climber, assisted by the ropes and the harness, can use both the grip provided by the special shoes on the smooth walls, and the cracks as a support point.
Multipitch is more like a set of different walls with some stations in the middle which allows the climbers to have a break before reaching the top.

[5] **Smearing:** it happens when the climber does not have a support in the rock and he/she has to use the shoe's rubber for friction against a lean wall.
[6] **Slabs:** it is a type of climbing where the rock have an angle less steep than vertical.
[7] **Cracks:** it is a style of climbing who deal with some open cracks in the rock and those allow climber to fit the foot in.
[8] **Mild overhangs:** it is a type of rock that have a steep angle slightly higher than vertical.
[9] **Multi-pitch:** it is a type of climbing who refers to the presence of multiple intermediate stations.

3.1.4 Shoe asymmetry

Final relevant aspect which would help buyers to understand the model based on their preferences is the asymmetry of the shoe.

During the act of climbing, the climbers look for the big toe support. Through this they are able to drive more force when some small rock edges comes into play. For this reasons high-level climbing shoes comes with this asymmetrical shape.

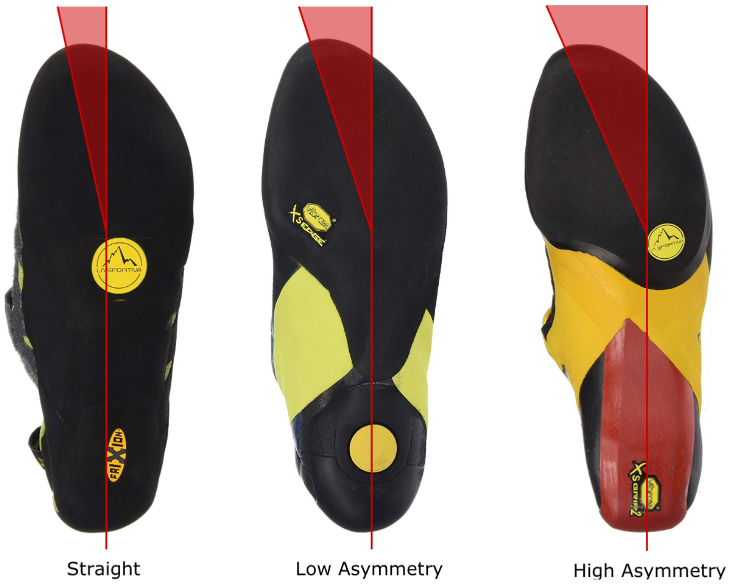

Figure 3.7: Anatomical structure of the lower side of a climbing shoe [8].

3.2 Downsizing issue

Until now they have been listed all the features necessary to buy a climbing shoe to fit specific conditions, such as the route that the climber wants to climb or the technique used. Though, one of the hardest variable to be predicted is the downsizing that a buyer would accept while climbing.

As said climber would expect to have a pair of climbing shoes being tight and enveloping. It can be obtained both by choosing a suitable model, but also by decreasing the size of the shoe, taking into consideration objective and subjective aspects.

3.2.1 Foot anatomy

Dealing with a new set of climbing shoe, the manufacturers themselves try to allocate their models of climbing shoe into a specific area, to direct new buyers to the right product. Two

relevant aspects that are covered from the brands are: foot shape and toe shape.
Foot shape is intended as the ratio between the length of the feet and the width of the feet. The length is computed and the hypothetical line that connect the big toe and the heel. The width is indeed considered as the hypothetical line that would connect the big toe and little toe along the foot knuckles.

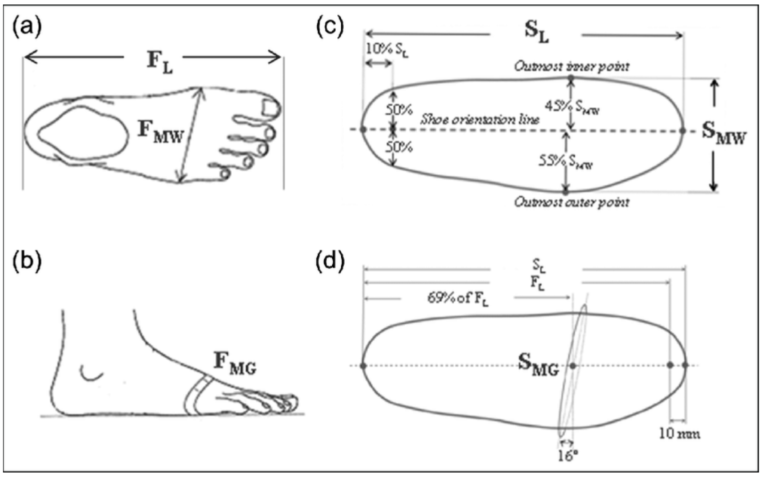

Figure 3.8: [1] Measurements of the foot and the inner-shoe dimensions. (a) Foot Length (FL) and Foot Metatarsal Width F_{MW}.

The second aspect relates with the toe shape. It is indeed possible to classify the toes shape mainly into three different clusters: Egyptian shape, Roman shape and Greek shape. The reason for this classification lies on the fact that each shape can be better with a specific model of shoe.

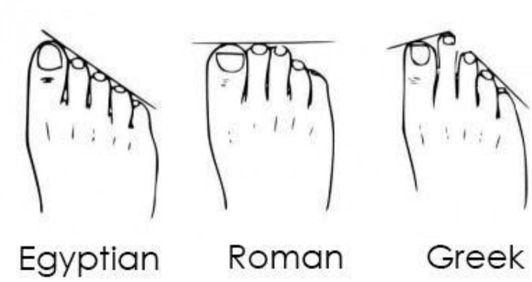

Figure 3.9: Shape of the toes more frequent in the population.

Let's assume that a new user owns a Greek toes shape while it was decided to buy a climbing shoe with a very tight toe rubber and an highly asymmetric shape. Greek toes shape are characterized for having the second toe as the longest. The anatomy of the feet would cause to the second toe to hit on the toe of the shoe and cause pain to the climber. To

overcome this issue the buyer needs either to the change model either to look for a bigger size.
The drawbacks on looking for a bigger size lies on a less enveloping climbing shoe, less feeling and confidence while climbing. Manufacturers generally suggests for a specific toe shape to be associated with a specific climbing shoe model. This is a symptom of how important it is to look for a shoe that can meet the anatomical needs of the climber.

3.2.2 Toes posture

As seen in the subsection 3.1.2 it was explained how the shoe is intended not only to envelop the foot, but also to constraint it. While climbing, the foot is stressed and positioned with unnatural poses. The sport man needs to trust that the shoes would provide the best grip while in those situations: the climber must feel confident that he can attach himself to a certain ledge. If the climbing shoe were not tight enough, the risk of slipping will increase and it will lead to dangerous injuries.

A climber has to consider three main areas to be tight.

1. The heel cup [Figure 3.2] is generally very tight and the anatomical shape of the feet needs to adapt this area, leaving no air bubbles while wearing the shoe. If so, the heel could slide and be very painful.

2. The second area relates with the last [Figure 3.2] which it deals with the width of the foot.

3. The edge [Figure 3.2].
   While wearing aggressive climbing shoes [Figure 3.6] climbers need to unload all the strength and the weight in a small overhang of rock.

To reach this tight feeling across all the foot it's necessary to look for climbing shoes with few sizes less than the standard shoe. That adjustment returns a good wrap on the heel and the middle of the foot. This choice tough provokes a severe toe compression.
The toe compression is a side effect of the downsizing, but it turns out to facilitate climbers. Moderate and aggressively downturned climbing shoes are intended to be used while climbing difficult and extreme walls. In fact, thanks to this curvature of the toes, the climber is

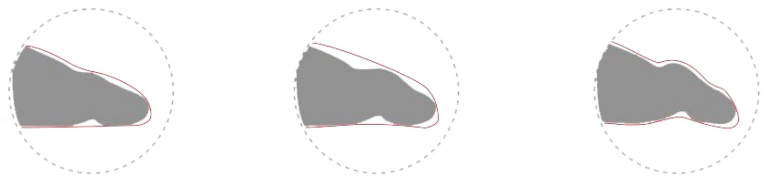

Figure 3.10: Toes posture for a neutral, moderate and aggressive downturned climbing shoe [3].

able to exert greater force on the overhangs. This implies greater performance in the climbing phase, and greater calm, since the shoes will provide feedback to the athlete on how much they can trust or not the grip of the shoes.

Although the shoe is constricting, it is not directly a consequence that a climber is able to compress the toes sufficiently. This posture is in fact very painful, and a callus developed on the knuckles of the toes helps to withstand the compression caused by the shoe. Climbers that are used to frequently climb are able to maintain this compression for a longer time. Since this compression became over the time an advantage, climbing shoe brands started to facilitate this posture.

3.2.3 Shoes materials

Climbing shoes brands created shoes with material such as lined leather or unlined leathers. Some other shoes are instead made with synthetic leather. This choice donate to the shoe a specific extensibility.
Lined or unlined leather shoes can stretch up more. Because it allows better extensibility, it allows climbers to resize the shoe more, since it can deform more.

Aggressive and moderately curved shoes are composed of these materials, since they allow the shoe to be deformed at the specific foot and it boosts climbers confidence. On the other side the neutral climbing shoes are more rigid and stiff, made with synthetic leather: they are comfier and they allow climbers to accomplish longer routes without suffering for the tightness of the shoe. Also the stiffer sole allows to use techniques as smearing [Figure 3.12][Figure 3.13], since the overall strength of the shoe would allow to exercise more strength all over the last [Figure 3.2].
It is worth a mentions that lined/unlined leather tends to loose more with respect to the synthetic one. This means that on a long term the lined/unlined leather shoe would result in a less tight shoe and it needs to be replaced with a new one. On the other side, synthetic

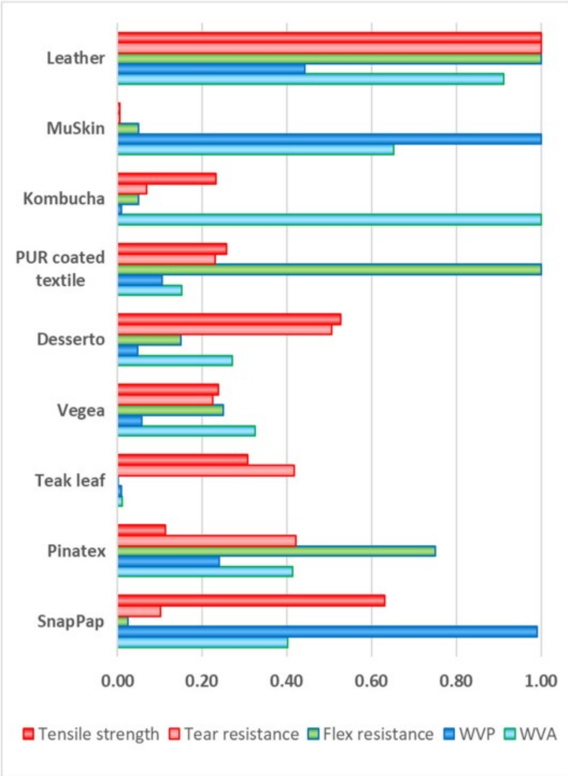

Figure 3.11: Features comparison between leather and synthetic leathers options [5].

Figure 3.12: Smearing technique.

Figure 3.13: Smearing technique.

one would instead allow a less initial downsizing due to the high resistance returned by the material, but it would preserve more the shape on a long term usage.

3.2.4 Brand differences

When it comes to buying a new model it is necessary to consider the different policies adopted by different brands. Brands in fact are known having a general fit such that it would be taken as a major guideline for every climber. It is common that the same person would fit good in two different models of climbing shoe while wearing different sizes. Although the models have the same characteristics for most of the structure, despite having the same shape and are built of the same material, this is not enough to define the size of the shoes

to buy.
It follows that these relationships between brands need to be taken into account. However, it becomes difficult to buy a climbing shoe model of a different brand from one already owned. Both because of the constructive and technical differences, and because of these specific resizing choices of the brand itself.

3.3 Physical issues

The curled position on the toes, the tight fit of the shoe and the squeezing shoes lead to some physical issues. The majority of the injuries are indeed a consequence of wearing climbing shoes with strict and unnatural constraint positions.
Major issues could be nail infections, pressure marks, big toes deviations neurologic complaints and subungual hematoma[10] [9]. Even if 80%−90% of the climbers reported to suffer while climbing, they accept it for improved performances.

Figure 3.14: Ray scans of a climber. A: Climber's bare foot standing; B: climber's foot wearing a climbing shoe; C: side view while wearing a climbing shoe [9].

Another common issue is related with the hallux valgus deformity. It was in fact clear that the constraint position would provoke an increasing possibility to encounter this pathology [Figure 3.14-B].

It is possible to underline that almost all climbers with 10 hours of climbing per weak on average, and with at least 5 years of experience, would report a 2.3 sizes downsizing on climbing shoes. Though all of them present a chronic condition.

[10]**Subungual hematoma:** it is the medical for bleeding under the toenail.

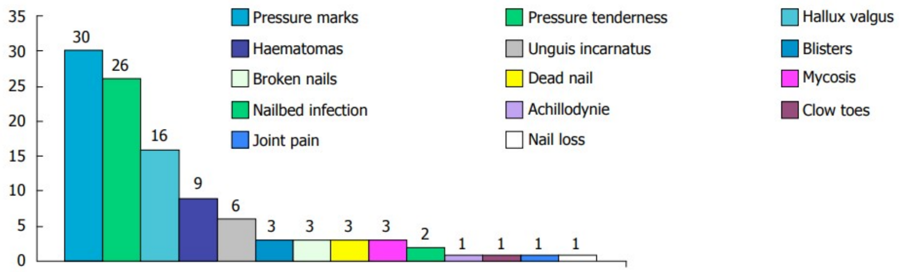

Figure 3.15: Average pathologies conditions on 30 male individuals with 5 years experience and 10 hours per week climbing [9].

Chapter 4

Data collection and analysis

As it was pointed out, it is demanding to deal a complex piece of technology as described in 2.3. The impediments that must be taken into account cannot be easily overcome. Beyond this a huge effort would not directly provide a good result in terms of quality of the predictive model. It is therefore necessary to proceed through a more gradual and logical methodology. It was decided to look for a baseline model that explains what the quality of a predictive model would be.

4.1 Data collection: methods and data

To take advantage of a more basic approach, it was necessary to have a certain amount of data and information from people who experience climbing. Said that the first step involved the collection of data.
To execute this process it was necessary to lean on a questionnaire service. It allowed to distribute the link that would allow the user to easily fulfill the questionnaire. The data returned from the platform are a Comma-Separated Values (CSV)[11] file which would help to analyze the information retrieved.
Data collected covered different areas, each one for an intended scope:

1. **User mapping**
2. **Core questions**
3. **User contacts**

[11] **CSV:** the comma separated values is a file format that is intended to import or expert some tables of data. Each row of text is separated with a separator character which would bound each column.

With the user mapping section it was possible to ask for the most basic questions to map the climber we are contacting. It contains information about the demographic data. It refers to user name, the age and the gender.
Together with those information it was asked to describe with a couple of questions about the physicality of the feet. The initial approach described in 2.3 was about a procedure about extracting those measure from a video uploaded from the user. This solution is certainly effective, but implementing it right away is a costly procedure. At first it was removed in favor of some additional questions in the questionnaire. To the user it was indeed asked to answer about:

1. Foot shape. The possible choices would converge into 4 different options: Narrow, Medium, Wide and Extra Wide. The user is asked to observe his own foot and to classify it in relation to one of the 4 available options.

2. Toe shape. As explained in 3.2.2 the position of the toes is very impactful, both in terms of performance in the climbing phase, but also in terms of pain while wearing the climbing shoe. Toe shape is another relevant features to be taken into account while buying climbing shoes. As seen in 3.2.1 the toe shape and the model used together are features that would determine the size of the climbing shoe, altogether with the actual feeling returned to the climber. The user is then asked to report the toe shape: Greek, Roman or Egyptian.

3. A sneaker size. To have a reference on the actual size of the feet instead of asking for a measurement it was asked to return the size a shoe daily used.

Finally, the system used to evaluate a climbing wall and the degree that the user feels confident he can climb on sight was also asked.

The core questions are describing all the climbing related information. In this case it was necessary to ask for all the information about the climbing shoes already owned and the feeling about wearing and using it. It was asked to the climber to add up to 5 different models of climbing shoes, together with the brand that is producing the shoe. Together with the model of climbing shoe already owned, it was asked to add the shoe size related to each model.

With those set of different information it is possible to map each specific model of shoe. This means that the questionnaire is highlighting if it was possible to extrapolate some spe-

Figure 4.1: Questionnaire page asking about the brand of the climbing shoe the climbers already own.

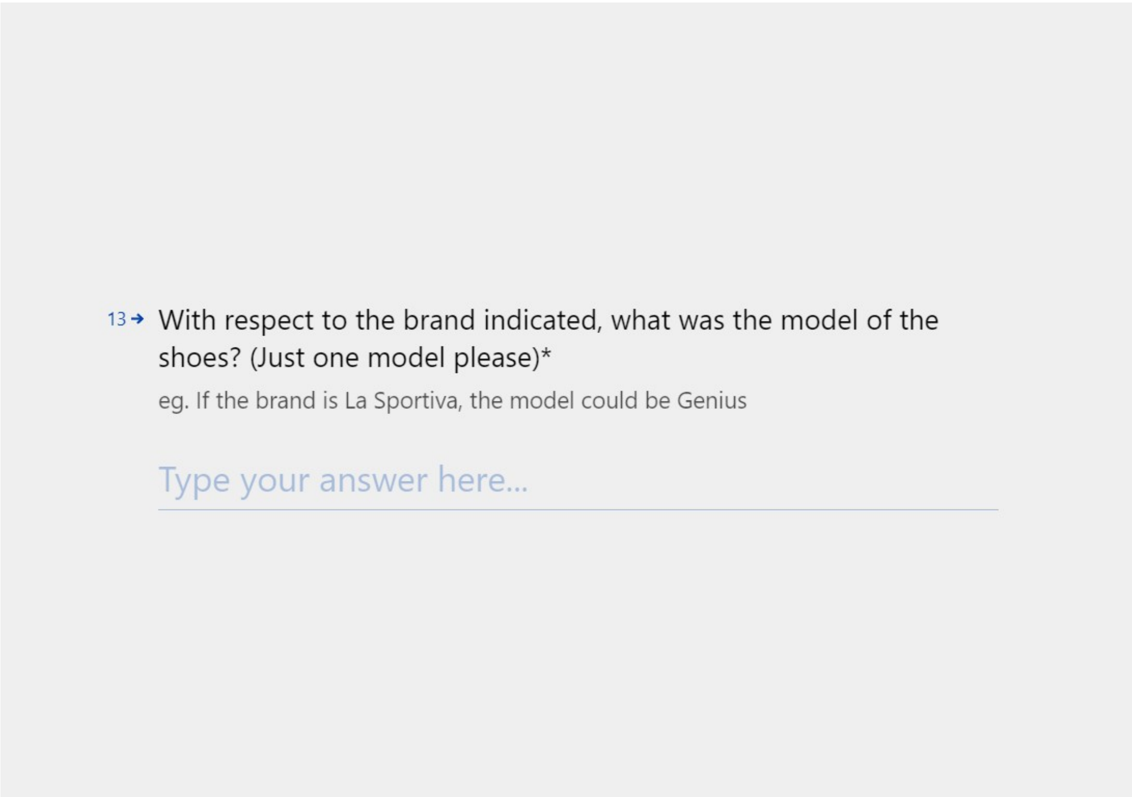

Figure 4.2: Questionnaire page asking about the model of the climbing shoe, related to the previously added brand, the climbers already own.

cific correlations between different models and brands.

Together with those information it was asked about the closure of the specific model. As

said while explaining the closure system in 3.1.1, it affects the climbing shoe downsizing. Models come with a specific type of closure and it was asked to associate to each inserted model of climbing shoe the closure that comes with that shoe.
Apart from this, as said in 3.1.3, each model of climbing shoes are built and distributed from the different manufacturers to be associated a specif style of climbing. Each model has been designed and conceived to perform better on a specific surface or steepness. Nevertheless it was necessary to have a feedback from the user side. It was indeed asked to the user to associate two different values to the inserted model of climbing shoe:

1. What kind of route the user uses the climbing shoe in. This means that climbers need to choose if the shoes had been use for: *Multipitch*, *Crag* or *Boulder*. This question is done to map under which general environment the user are wearing the climbing shoe.

2. For which climbing style those shoes are performing better. In this case it was asked to understand under which specific type of wall, steepness or condition this pair of climbing shoe is performing better: *Slabs*, *Vertical*, *Overhangs*, *Bouldering*, *Resistance* or *All styles of climbing*.

Although the shoes are assumed to be bought following the instructions of the manufacturers, it is not something it was to be taken for granted. Climbers, mostly experts, could be more sensible and look for more specific characteristics of the model itself for a specific route. Since that it was necessary to have a more direct feedback from climbers to map in which cases a specific model or brand, with those specific features are more likely to perform better.

Last topic covered with the Core Questions relates with the comfort level felt while wearing that specific model. As explored during 3.3 there are more than one reason to be aware of the feeling while wearing climbing shoes. Even though climbers are aware that an uncomfortable wearing means performance however it should be remembered that it could lead to some irreversible damages.
It was asked to the user to evaluate on a scale from 1 to 5:

1. The pressure on the knuckles of the toes. A toe tighten curled of the shoes would provide a higher performance while climbing but also a very painful feeling on a long term climbing sessions. An excessive pressure on the knuckles would highlight how the climbing shoe size is not perfectly fitting the user foot.

2. Pain on the Achilles Tendon. The pain on the Achilles tendon is related with a not completely enveloping heel cup [Figure 3.2]. Under this assumption there's nothing that a climber can manage to do. Climbing shoes are made with a specific heel cup: if the buyer's heel does not fit perfectly, all you can do is change the model.

3. Overall comfort while climbing. It is a measure of how good is the experience with the actual pair of climbing shoe.

In the main question it was decided to ask if the climber uses socks when climbing. It is generally not in common use, however some climbers have reported that for comfort and convenience they are used to wearing them. In fact, you can slip on your climbing shoes more easily. In addition, the pain felt while wearing the shoe is less. As well as the sweat that could lead to irritation and slipping inside the climbing shoe is much less. However, there are also several negative sides: a lower sensitivity, a less enveloping shoe and the need to adapt the size of the shoe to the use of the sock.
The last section is made to let the user accept the privacy and policy. In the end, an e-mail was requested and the willingness to collaborate in the future for the development of the project itself.

At the end of the process it was possible to return more than 2500 answers to the questionnaire.
On the other side, the vast majority of the people who compiled the questionnaire owned at least 2 models of climbing shoes. This means that it was possible to collect a greater number of reviews related to climbing shoes, as many users entered more reviews in the same questionnaire.

4.2 Data Analysis

4.2.1 Correlation hypothesis and analysis

The set of data collected allowed to examine the results.
Before looking for any specific relationship between data, it is necessary to identify which relationships are plausible and need to be searched.
Climbing experts are aware of the constraints they take into account when buying a new pair of climbing shoes. It has been decided to interview a small group of climbers who answered our questionnaire previously, asking for any feedback while testing a new set of

climbing shoe. Therefore, a set of observations have been defined that would coincide with the conditions regarding the reliability of a size for climbing shoes.

The climbers interviewed included both experts and amateurs. Users who had always and only purchased in physical stores for years and users who had already had experiences with purchases of climbing shoes online. The relationships found were identified as being shared by the majority of respondents.
The relationships that were analyzed were to evaluate whether there was a relationship with respect to downsizing. Some of those were:

1. Frequency of climbing
2. Expertise
3. Feet shape
4. Toe shape
5. Climbing shoe closure
6. Usage of socks while climbing
7. Comfort

The frequency with which a climber is used to practice was every time mentioned as one of the first attributes to own to downsize more the climbing shoe. While climbing and continuously adding some pressure to the knuckles, the climber is able to generate some callus. Those would allow climber to bear more a tighter shoe. It come straightforwardly that callus would decrease the amount of pain felt by the climber. A climber used to practice almost everyday is able to diminish the size of the climbing shoe and to resist more a tighter shoe with respect to who practice less.

Speaking about the expertise climbers highlighted how an expert climber would be able to be more aware about how to wear a set of climbing shoe. In this case it was evident that only expert climbers underlined how experience had helped them to understand that certain types of shoes require a certain posture inside them to be worn. The amateurs were unaware that the more aggressive climbing shows needed to be bought with several sizes less than the sneaker shoe size.
Leaving that aside, an expert all over the years deformed their own feet. It means that their

feet are more likely to accept climbing shoes shape.

As already explained in 4.2, the shape of the feet and of the toes are relevant while a climber should deal with downsizing.

For what concerns closures about climbing shoes, climbers with more experience reported that the three different type of closures would affect the size of the climbing shoe. Slip-on climbing shoes tend to relax more over time. It is therefore necessary to take into account that over time they will become less enveloping than a climbing shoe with laces: the laces can in fact partially compensate for the sagging of the climbing shoe materials.

Regarding the use of socks, it was a logical reasoning regarding the fact that the use of socks requires to downsize less the climbing shoes.

Finally, the investigation into comfort was a way to define if those who were used to buying a narrower size of shoes reported having a less comfortable fit.

4.2.2 Results

At that point the data gathered from the questionnaires were analyzed and processed to verify if any of those relations where immediately verified.

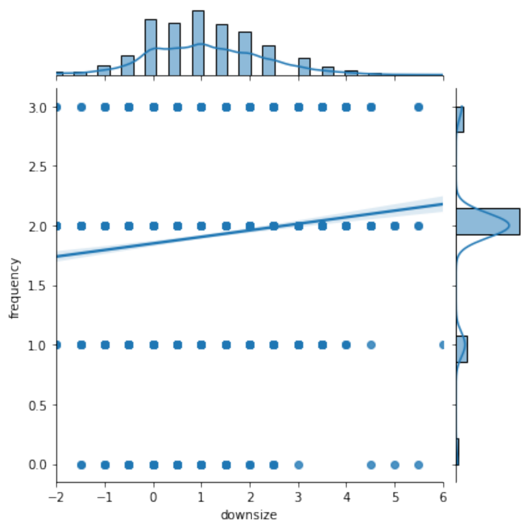

Figure 4.3: Relations between the frequency and the downsizing.

The figures [Figure 4.3][Figure 4.4][Figure 4.5][Figure 4.6] are highlighting along the axis the distribution of each value. Along the y axis it is described how much a climber downsized that specific model. In the figure [Figure 4.3] along the x axis it is described the frequency a climber is used to climb. Each value was associated to an integer value: 0, 1, 2, 3.

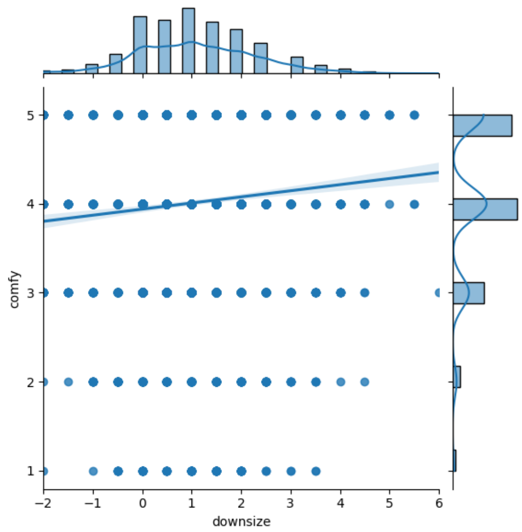

Figure 4.4: Relations between the comfort while wearing the climbing shoe and the downsizing.

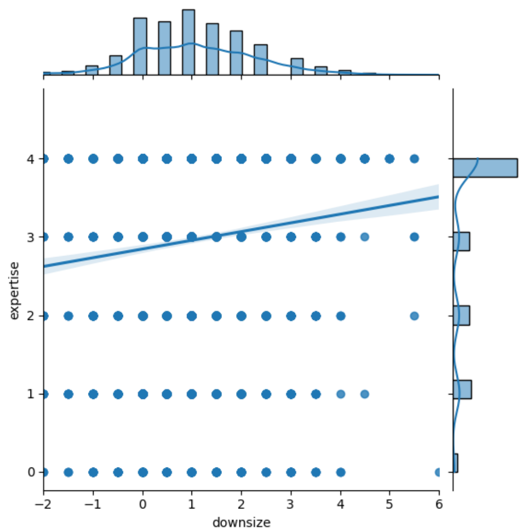

Figure 4.5: Relations between the expertise of the climber and the downsizing.

In this case:

- **0**: less than once a week
- **1**: once a week
- **2**: 2-3 times a week
- **3**: everyday

What it was expected is that the downsizing would increase together with the frequency with which the climber is used to climb. Observing the regression line described from

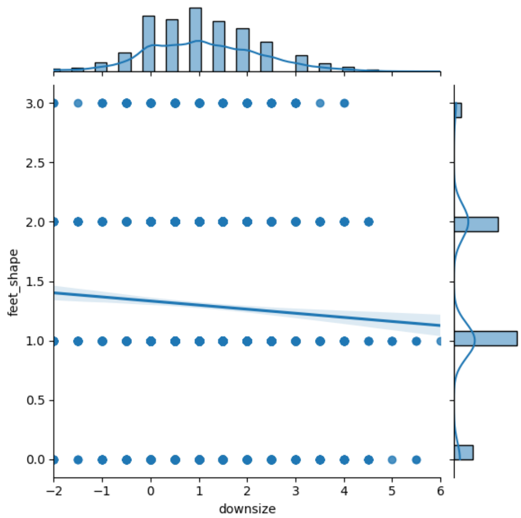

Figure 4.6: Relations between the shape of the feet and the downsizing.

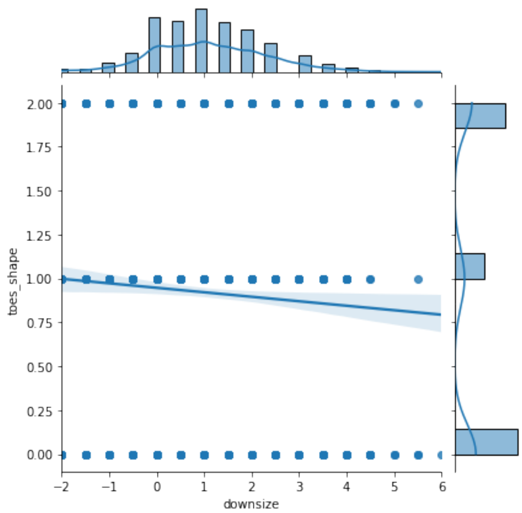

Figure 4.7: Relations between the shape of the toes and the downsizing.

the defined function of the Seaborn[12] library "jointplot", a climber that practice more frequently would downsize more the climbing shoes.

In the figure [Figure 4.4] it was expected that an increasing comfort would coincide with a lower downsizing, meaning that tight shoes would have been less comfortable. On the other side what the users returned is that generally whoever is wearing a smaller size in climbing shoes is reporting to have a more comfortable fit with respect to the others.

The figure [Figure 4.5] is representing the relationship with respect to the expertise. Also in this case it was converted the expertise values into some integers values:

- **0**: less than six months of climbing

[12]**Seaborn:** Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

- **1**: one year of climbing
- **2**: 2 years of climbing
- **3**: 3 years of climbing
- **4**: more than 3 years of climbing

It was expected to see that an expert would downsize more the size of the climbing shoes. In this case it is visible that the regressive model is highlighting this behaviour.

The figure [Figure 4.6] relates with the shape of the feet. The same procedure was applied here:

- **0**: Narrow
- **1**: Medium
- **2**: Wide
- **3** Extra Wide

It was expected that a Narrower feet dimension would return a higher downsizing while buying climbing shoe. It is indeed true that whoever compiled the questionnaire with a narrower feet generally would downsize more.

Last figure [Figure 4.7] refer to the relation between the shape of the toes and the downsizing. As in the previous case it was used a conversion such that:

- **0**: Egyptian
- **1**: Roman
- **2**: Greek

The Greek toes do have a longer second finger. It means that a climber would expect to downsize more to have a better feeling in the big toe, in exchange with an increasing pressure in this second toe. In this case the graph is not highlighting a very precise relation between the toes shape and the downsizing procedure.

Those relationships seems to partially confirm the analysis. Even though the automatic regressive model seems to be reliable and agree with the theoretical hypothesis, there is not a sharp relation between the any of those feature and the downsizing procedure.
It is indeed visible that the distribution of the answers present a high variance and the distribution of how climbers are used to downsize are not bonded on one of those features.

4.2.3 Analysis of the results

The results of the analysis of the data has highlighted that the assumptions made can be confirmed from the users who took the questionnaire. Left aside the relation between the shape of the toes and the downsizing procedure, climbers downsizing behaviours are aligned with the hypothesis previously made.
On the other side the outcomes of the downsizing procedure seems to have a high variance[13]. It does mean that the downsizing values are distributing as if it was a Gaussian curve with with a high dispersion value from the actual expected value. Said that, it was complicated to understand if any of those relations were actually sufficient to cluster a new climber in a specific downsizing behaviour. Analyzing the figure [Figure 4.5] the downsizing for whoever is climbing since more than 3 years is covering all the range of the outcomes. It was possible to find more than one expert climber answered that he/she used to wear a pair of climbing shoe with a bigger size with respect to the actual sneaker shoe size.

In order to create order and to have guidelines, some specific conclusions have been defined for the previous analysis:

1. it had been verified that there exist a weak direct correlations between climber habits and choice for the right size

2. it even proved that some anatomical features, such as the size of feet, could determine an alteration when it comes to deciding which shoe size to prefer between one and the other

3. it had been seen that it is possible that a set of different incoming features would provide a solution for a new climber which could converge on a single downsizing value

4. it has also been shown that it is not possible through the exploit a single feature to restrict the possible values in the prediction phase

.

[13]**Variance:** it is a measure of dispersion, meaning it is a measure of how far a set of numbers is spread out from their average value.

4.3 Climbing shoe models analysis

The results of the previous analysis led to the search for further relationships within the collected data. More specifically, as already noted in Chapter 3, the individual shoes are a constraint for the choice of shoe size.

It was therefore decided to shift attention along this direction, trying to understand if it was possible to extract binding information from the single model of climbing shoes. It was possible to take a look and identify the submitted brands of climbing shoes, cluster the data based on a single brand and search for any specific model.

Looking for this information we filtered the entries in the dataset and it was possible to list the major brands available:

- La Sportiva
- SCARPA
- Five Ten
- Tenaya
- Evolv
- Ocun
- Mad rock
- Unparallel

For each of those brands it was possible to distinguish for a vast amount of different models. The largest number of added reviews related to climbing shoe models are related to:

- "Solution" from La Sportiva brand with 5.45% of the entries
- "Skwama" from La Sportiva brand with 5.22% of the entries
- "Instinct VS" from SCARPA brand with 2.89% of the entries
- "Drago: from SCARPA brand with 2.89% of the entries
- "Theory" from La Sportiva brand with 2.58% of the entries

and so on. Around 70.6% of the entries are related with La Sportiva and SCARPA brands. It was decided to focus the research focusing to those brands to search if it was possible to extract any information about a downsizing alteration between different models.

The analysis started with two diametrically opposite pair of climbing shoes. It was decided to analyze for an aggressive pair of climbing shoes and a pair of neutral climbing shoes. The relationships between who wore this model of climbing shoes were available, as well as information about the experience and frequency with which the climber usually climbs. It was supposed that an aggressive climbing shoe would accommodate a generally higher downsizing for what concerns climbing shoes.

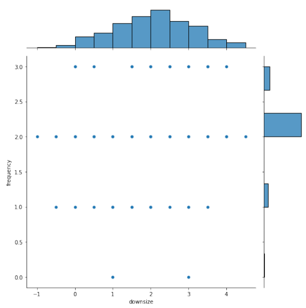

Figure 4.8: Distribution of downsizing for the Solution model of La Sportiva climbing shoe.

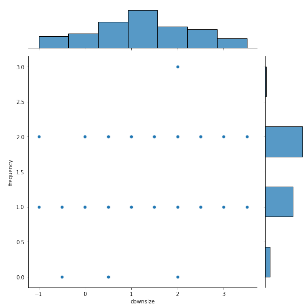

Figure 4.9: Distribution of downsizing for the Tarantula model of La Sportiva climbing shoe.

The downsizing distribution for the models for the Tarantula climbing shoes highlights:

- Mean: 1.237
- Variance: 1.18
- Maximum downsize value: 3.5

while the Solution climbing shoe distributions highlighted:

- Mean: 1.8
- Variance: 1.07
- Maximum downsize value: 4.5

The mean value underlines that the aggressive climbing shoe (Solution, [Figure 4.8]) is appreciated with a more extreme downsizing with respect to the neutral one (Tarantula, [Figure 4.9]). Also the maximum downsize values reported that a very dramatic diminishing of the size is viable with the aggressive model, while the neutral one didn't meet any review.

An even more interesting observation comes from the relation between the downsize and the climbing grade reported from the users. As in the previous case the intention is to identify some kind of correlation between a specific model shoe and a specific category of users.

As in the previous case the data were filtered considering the two different models of climbing shoe. Together it was applied the Gaussian Mixture Model (GMM) considering 3 different Gaussian components for what concern the grade distribution. It is expected that it is possible to distinguish between three cluster of different climbers based on the capabilities. Whoever is comfier climbing at a higher level is expected to wear a tighter pair of climbing shoe.
During the process of reviewing a pair of climbing shoe, the user is asked to report the grade with which is confident to deal with. Nevertheless it does exist different scales to measure the level of difficulty of a climbing wall. It is then necessary to normalize and make the scales comparable.

4.3.1 Climbing grades and scales

As said in section 4.2 it was for the system and grade a climber can climb on sight.
It was possible to select from a different set of systems from the questionnaire:

- French
- USA (YDS)
- Australian/NZ (Ewbanks)
- UIAA[14]
- South African
- British
- Other

From this set of possible systems it came out that another system was vastly used from a lot of climbers, that's V-Scale[15].
It was decided to set the wider used as a reference scale: the French scale. Once set it as a

[14]**UIAA :** Union Internationale des Associations D'Alpinisme, It is an association that in the 1967 tried to define and build a new scale, called *"Scale UIAA"* to gather together all the others.
[15]**V-Scale:** the V-scale, also called Hueco "V" grades is the most widely used system in North America.

reference it was possible to cast each of the other value for each scale to the French scale.

4.3.2 Results

From the actual analysis of the downsize procedure and the grade of the climber the actual result is showed in the images [Figure 4.10][Figure 4.11].

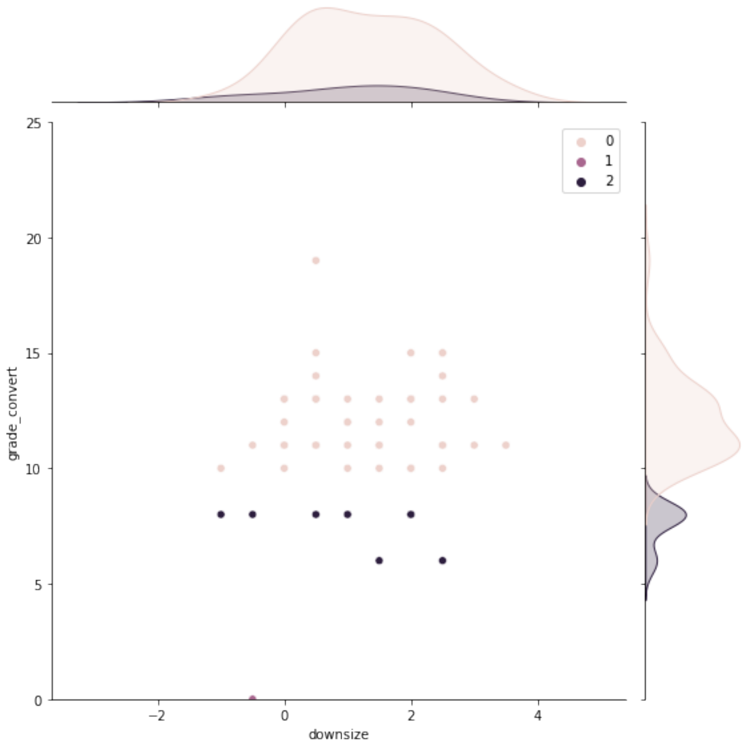

Figure 4.10: Distribution of downsizing for the Tarantula model of La Sportiva climbing shoe.

The Gaussian Mixture Model should be able to individuate and divide the climbers into grades based on the capability.
And indeed it is, since for what concern the Solution climbing shoe [Figure 4.11] the three labels returned from the GMM highlights this behaviour. Speaking of the Tarantula, the [Figure 4.10] is reporting mainly 2 labels. These labels refer to medium and basic climbing walls. It indeed seems that the higher grades component is missing.

The y-axis related to downsizing would refer to the downsizing. In the [Figure 4.11] more than the other, it is visible that the separation due to the GMM, split the downsizing distribution in three curves with a different mean value, that meet the expectations. The peak of the curves show that the harder is the wall that a climber is used to climb, the tighter the shoes would be. And vice versa the easier the wall the looser would be fit on the climbing

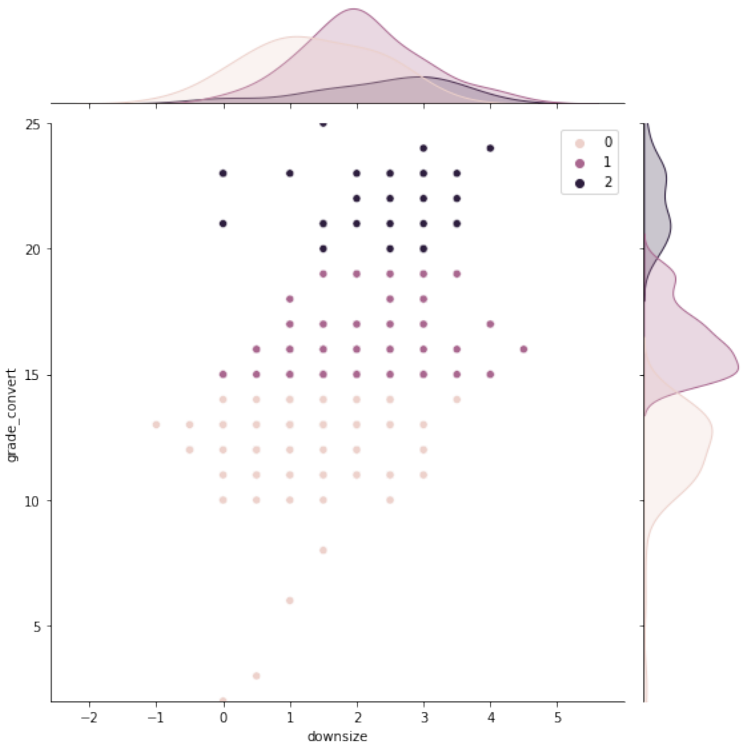

Figure 4.11: Distribution of downsizing for the Solution model of La Sportiva climbing shoe.

shoe.
So, it is possible to assert that relation between the grade and the downsizing seems to bring some information.

All the information extracted from this analysis leave now space to a wider and deeper process of investigation and application of procedure and techniques that would converge to create a prediction model. This would be topic for the next chapter, that would explore the approaches and the solution adopted.

Chapter 5

Predictive model

Along this chapter it will be explained the set of different techniques used to solve the prediction task, starting from the exploratory solution to the implemented one. During the process it has been decided to look for white-box predictive models. White-box approaches are characterized by a lower predictive capacity. A white-box approach, such as a linear regression or a decision tree, are in fact not able to model the intrinsic complexity present in the dataset. The biggest advantage is that you are able to interpret the predictor's behavior. A black-box approach which, on the other hand, is more performing, but makes predictions that can be corrected without having total control of the model used. The predictive models presented are indeed two white-box approaches. Finally, a plausible future approach has been added.

5.1 Content Based and Collaborative Filtering approaches

5.1.1 Collaborative based filtering

The first approach saw the use of Collaborative Filtering (CF) for a Recommendation System.
The collaborative filtering approach exploits the evaluations on items to create an adaptive web. It takes the opinions of the customers and it creates a large interconnected community made by reviews on the full set of items.

The main concept of this piece of technology involves a process of classification with an evaluation that goes to indicate the degree of satisfaction for that particular object. It is widely applicable on websites. One of the most famous refers to a website that is based on movie reviews. These are ranked a value from 1 to 5 based on the degree of appreciation of

the user. If the two users share a wide set of opinions about movies, a new film appreciated by one of the two would probably meet the tastes of the other.
In that way it is possible to create an association between the user and the item. This association creates a matrix such that along the column it is possible to see the different users, while along the rows the connection strength between the user and the item itself.

To fit this technique to our scope, it is necessary to identify a value that associate a satisfaction value between the user and the climbing shoe it is using.
The specific field of the questionnaire that would meet our scope is the evaluation about the comfort level about that specific model of climbing shoe. In the scenario of buying a new pair of climbing shoe, it could be interesting to log the online website, to provide a set of information about the climbing shoe, and receive a suggestion about other models that encounter a good review from people with similar opinion about the same models.

5.1.2 Content based approach

Simultaneously to the Collaborative Filtering it has been implemented a Content-based approach for a Recommendation System.
Even in that this approach fits with the creation of an adaptive web: it aims to provide a personalized suggestion to the user. Differently from the CF, the content-based approach aims to discover elements that are more similar to each other, taking into account a large set of different characteristics. Reusing the same example as before in the field of Recommendation System, this type of approach would suggest to user a new film to be seen based on the features of the two films. It extrapolates the recurring characteristics of the films that the user usually sees. It groups them, and suggests new films to be seen only if the latter have the same distinctive features of the films appreciated by the user.

Even in that case, it had been possible to adapt this approach to our solution.
As previously said, the vast majority of the incoming data would cover the La Sportiva climbing shoe brand. It has been decided then to implement this approach for what concern the models available of that brand. It has been possible to create a dataset containing all the specifications about the climbing shoe models of La Sportiva.
The set of different information extracted for each type of model, is then used to map each model closer to other models that own similar features [Figure 5.1]. A climber that is enjoying a pair of climbing shoe would be suggested to buy a pair of climbing shoe that own similar features.

Shoe	Construction	Downturn	Level	Use	Environment	Closure	Material	Stiffness	Fit	Midsole	Technology	Last Shape	Top	Lining	Thickness	Weight	Foot width		Notes
Theory	Slip lasted	Aggressive	Intermediate Advanced	Bouldering Face	Indoor	Velcro	Leather	NA	Performance	Without midsole	No-edge Vibram	Asymmetric	Low	Unlined	4 mm	200g men/ 180g women	Medium Wide		
Solution	Slip lasted	Aggressive	Intermediate	Sport Overhang Face	Inddor Outdoor	Velcro	Leather	NA	Performance	Partial midsole	Leather Vibram	Asymmetric	Low	Lined	3.5 mm	258g men/ 205g women	Narrow Medium		
Katana Lace	Lace-to-toe	Moderate	Intermediate	Sport Crack Overhang Face All-Around	Indoor Outdoor	Lace	Leather	NA	Performance	Full midsole	Leather Vibram	Asymmetric	Low	Lined	4 mm	247g	Narrow Medium		
Kubo	Slip lasted	Moderate	Beginner Intermediate	Bouldering Sport All-around	Indoor	Velcro	Leather Mesh Upper Synthetic	NA	Tech	Full midsole	Vibram	Asymmetric	Low	Unlined	4 mm	200g men/ 170g women	Medium		
Zenit	Slip lasted	Moderate	Beginner Intermediate	Bouldering	Indoor Outdoor	Velcro	Fabric/ Textile Knit Upper	Medium-Stiff (2/3)	Tech	With modsole	Vibram	Asymmetric	Low	Unlined	5 mm	240g men/ 205g women	Medium		
Speedster	Slip lasted	Aggressive	Intermediate Advanced	Bouldering Sport Overhang	Indoor Outdoor	Slip-on	Leather Synthetic	Medium-Stiff (2/3)	Performance	Without midsole	No-edge Vibram	Asymmetric	Low	Unlined	3 mm	190g	Wide		
Geckogym	Slip lasted Vegan	Neutral	Beginner	Bouldering Face	Indoor	Velcro	Synthetic	Medium-Stiff (2/3)	Comfort	Full midsole	NA	Straight	Low	Lined	5 mm	235g men/ 205g women	Medium Wide		
Tarantula	Slip lasted	Neutral	Beginner Intermediate	Sport Crack Overhang Face All-Around	Indoor Outdoor	Velcro	Leather	Stiff (3/3)	Comfort	With midsole	NA	Asymmetric	Low	Unlined	5 mm	290g men/ 240g women	Medium Wide		
Tarantulace	Slip lasted Lace-to-toe	Neutral	Beginner Intermediate	Sport Crack Overhang Face All-Around	Indoor Outdoor	Lace	Leather	Stiff (3/3)	Comfort	With midsole	NA	Asymmetric	Low	Unlined	5 mm	260g men/ 218g women	Medium		
Finale	Slip lasted Lace-to-toe	Neutral	Beginner	Bouldering	Indoor Outdoor	Lace or Velcro depending	Leather Synthetic	Medium-Stiff (2/3)	Tech	With midsole	Vibram	Straight	Low	Unlined	5 mm	226g men/ 221g women	Narrow Medium		
Mythos	Lace-to-toe	Neutral	Beginner Intermediate Advanced	Bouldering Sport Overhang	Outdoor	Laces	Leather	NA	Tech	Full midsole	Vibram	Asymmetric	Low	Unlined	4 mm	248g men/ 235g women	Medium		
Miura	Slip lasted Lace-to-toe	Aggressive	Intermediate	Bouldering Slab Sport Trad Crack Overhang Face All-around	Indoor Outdoor	Laces	Leather	Medium-Stiff (2/3)	Performance	Partial midsole	Vibram	Asymmetric	Low	Lined	4 mm	239g men/ 223g women	Narrow		
Skwama	Slip lasted	Aggressive	Intermediate	Bouldering Slab Sport Overhang Face	Indoor Outdoor	Velcro	Leather	Soft (1/3)	Performance	With midsole	Vibram	Asymmetric	Low	Unlined	4 mm	210g men/ 171g women	Medium Wide		
TC Pro	Slip lasted Lace-to-toe	Moderate	Intermediate	Trad	Outdoor	Laces	Leather	Stiff (3/3)	Performance	With midsole	NA	Asymmetric	Mid	Unlined	4 mm	247g	Medium Wide		
Otaki	Slip lasted	Aggressive	Intermediate Advanced	Sport Overhang Face	Indoor Outdoor	Velcro	Leather Synthetic	Medium-Stiff (2/3)	Performance	With midsole	Vibram	Asymmetric	Low	Lined	4 mm	218g men/ 175g women	Medium		
Futura	Slip lasted	Aggressive	Intermediate Advanced	Bouldering Sport Trad Crack Overhang Face	Indoor Outdoor	Velcro	Synthetic	Medium-Stiff (2/3)	Performance	Partial midsole	No-edge	Asymmetric	Low	Unlined	3 mm	225g men/ 196g women	Narrow Medium		
Miura VS	Slip lasted	Aggressive	Intermediate Advanced	Bouldering Sport Overhang Face	Indoor Outdoor	Velcro	Leather	Medium-Stiff (2/3)	Performance	Partial midsole	Leather Vibram	Asymmetric	Low	Lined	4 mm	270g men/ 228g women	Wide		
Oxygym	Slip lasted	Neutral	Beginner Intermediate	Sport	Indoor Outdoor	Velcro	Synthetic	Medium-Stiff (2/3)	Comfort	With midsole	NA	Asymmetric	Low	Lined	5 mm	202g men/168g women	Medium		

Figure 5.1: Database containing the models and the specifications for La Sportiva [4].

The measure of the similarity comes from an analysis due to the Word Vectorization, the Term frequency–inverse document frequency (TF-IDF) and the Cosine Similarity.

With the first step, the word Vectorization, it is possible to create a set of dense vectors that represents the words projection in the vector space [10]. The TF-IDF is a statistics that returns the grade of relevance of a specific work inside a document. In other words the more frequent is a word in a document, the higher is the importance of that word. In our case the Tf-IDF is asserting if a specific feature is present or not.

The cosine similarity, defined as

$$S_C(A,B) = \frac{A \cdot B}{||A|| \cdot ||B||} = cos(\theta) \tag{5.1}$$

represents the actual distance between represented vectors. Since our words representation now comes as a vector, it is possible to compute the angle between those. This angle, $cos(\theta)$ coincide with the Cosine Similarity: if $cos(\theta) = 1$, the two vectors are equal. If, otherwise, it equals to 0, it means that $\theta = 90°$ and that the two vectors are orthogonal.

It is then possible to plot the cosine similarity matrix.

As said, the closer the cosine similarity value is to one, the higher the similarity between

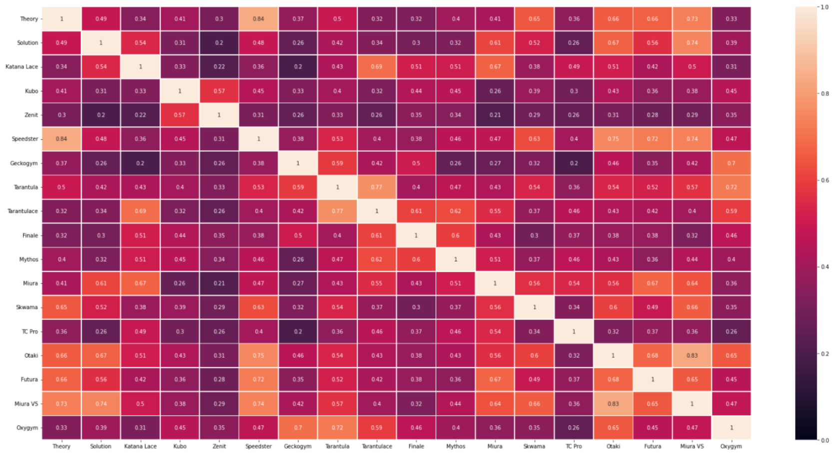

Figure 5.2: Cosine similarity matrix for La Sportiva brand.

two climbing shoes.

5.1.3 Results and limitations

The previous approaches differ slightly from what is the ultimate goal of e-commerce, linked to suggesting a climbing shoe number that is as suitable as possible for a new user. However, it must be considered that these two solutions remain a solid basis if we want to provide, together with the prediction of the number of the climbing shoe, a model with characteristics similar to one already owned. Furthermore, it should be considered that in the world of climbing shoes, the transition from one type of shoe to another is not easy, especially when passing from a more neutral model to a more aggressive model. This solution is a valid solution to identify which model the user can go to buy alternatively to his already owned model.

On the other side to fulfill this type of approach it would be necessary to explore different brands. Since brands use different technologies, rubber and materials to build climbing shoe models, it is not so immediate to have a suggestion that you switch from one brand to another.

Furthermore, this approach requires that the user who is purchasing the climbing shoe model already has a model owned.

5.2 Curve fit approach

During the collection of the data and the creation of the questionnaire, information emerged that describes the behavior of climbers during the purchase phase. Once the development

phase began, it was decided to follow this path in a more technical way, analyzing and creating mathematical models from previous observations. These mathematical models can in fact be used to exploit the Curve Fitting approach.
This approach finds the parameters that best fits the set of observations given a generic curve.

Most of the inputs values present in our dataset though has been one-hot encoded. The process of hot encoding derives from the necessity to cast a categorical label to a numerical one. The one-hot encoding create a set of different columns on the dataset starting from the set of categorical values of a feature. Each column would have a 0 or 1 value based on the presence or not of that categorical feature in that position.

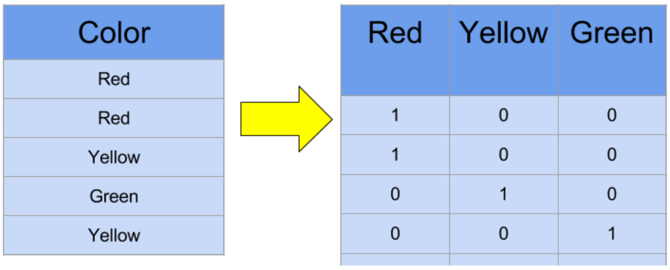

Figure 5.3: One hot encoding example.

The reason why one-hot encoding must be used is linked to the fact that some predictive models do not accept categorical variables. It is therefore necessary to go through a process encoding the variables to make them interpretable by the model.

Resuming the discussion about the curve fit, suppose to be interested to search for the best first degree order line that better approximate a set of data. The type of curve it is looking for is:

$$y = a \cdot f(x) + b \tag{5.2}$$

where a and b are define as a couple of parameters of the mapping functions, x is the input parameter and y is the calculated output. While a and b are defined as coefficients or parameters, x is defined as variable. The function $f(x)$ could describe different types of functions, as it will be seen in the equations 5.3. The equation 5.4 express all the possible lines while varying the coefficients.
The curve represented could vary based on the necessity:

$$y = ax^n + b$$
$$y = a \cdot e^{c \cdot x} + b \quad (5.3)$$
$$y = a \cdot \sin(x) + b$$

The reason because it has been decided to look for this approach comes from the assumptions that came from 4.2. Previously those assumptions had been made to investigate about the direct correlation between two specific features from the dataset. In this case it has been analysed he curve that could describe the relationship between the features given from the dataset. Since this information is the free result of the previous study, the effort required to test the curve fit was minimal.

Once it has been possible to describe the downsize value as a curve based on the input, it is possible to define a curve fit function that takes as input different values from the incoming data.

The function would present a set of n different coefficients:

$$y = a_1 \cdot f_1(x) + a_2 \cdot f_2(x) + a_3 \cdot f_3(x) + _{...} + a_n \cdot f_n(x) + b \quad (5.4)$$

where $a_1, a_2, a_{3,...}, a_n, b$ are the set of coefficients to be determined and $f_1, f_2, f_{3,...}, f_n, b$ are the functions that describes the curves for the relations between y and x.

Let's suppose to investigate about how the frequency on climbing could influence the downsize while buying a pair of climbing shoe.

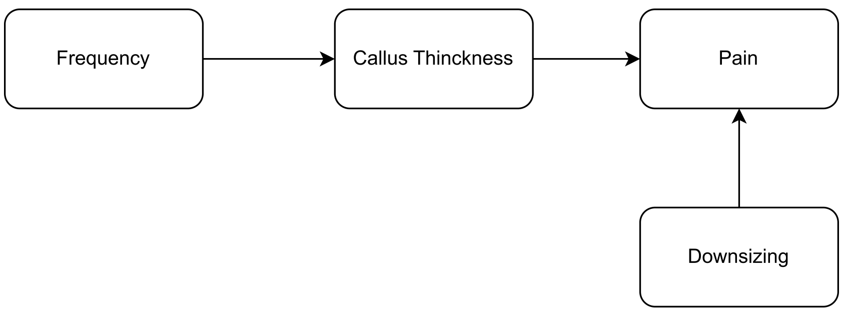

Figure 5.4: Diagram representing a possible relationship between the frequency and the downsizing.

Let's assume that:

- x = frequency, known variable

- y = callus thickness
- γ = pain on the knuckles of the feet
- s = downsizing

It is then possible to describe how each variable relates with the other.
It is known that the callus would rise more if the climber is used to frequently practice. It is possible to assume that it exists a linear relation between the two functions.
On the other side the pain that a climber would feel while climbing on the knuckles diminish if it is present a thicker callus. Nevertheless it was reported that it does not exists a very sharp difference once the callus is present. This behaviour could be described with an inverse exponential relation.
Finally the downsize value relates to pain as a linear function, since the less the pain, the more the user would downsize.
This reasoning has been used to describe all the subsequent relation that are present in this chapter. It has been created a set of $f(x)$ functions to define a relation such as the one presented in 5.4.
What we do end up with is:

$$\begin{aligned} y &= c_1 \cdot x + c_2 \\ \gamma &= c_3 \cdot e^{y} + c_4 \\ s &= c_5 \cdot \gamma + c_6 \end{aligned} \tag{5.5}$$

It is then possible to identify the value s that specify the downsize of the climbing shoe, starting from the x value with which it is possible to define the frequency with which the user is used to climb.
The outcome could be described as:

$$s = c_1 \cdot (c_2 \cdot e^{c_3' \cdot y} + c_4) + C_5 = c_1' \cdot (e^{c_2' \cdot x + c_3'}) + c_4' \tag{5.6}$$

and simplifying that it is possible to define:

$$s = c_1 \cdot e^{c_2 \cdot x} + c_3 \tag{5.7}$$

This procedure allowed to define a curve that could describe the state for the downsizing while varying the frequency value.

Similarly the same analysis had been done with other variables, such as the experience.

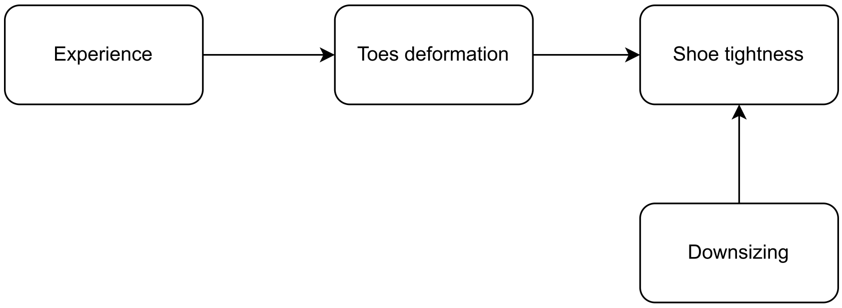

Figure 5.5: Diagram representing a possible relationship between the expertise and the downsizing.

Even in that case the experience linearly relates with the toes deformation, that are a direct consequence of the shoe tightness. As said in 3.3 the toes would deform passing the years wearing climbing shoes. This would imply an alteration in the choice of the climbing shoes, as already said. It follows that this relation is described from a linear equation between the experience and the downsizing.

Similarly while dealing with the width of the feet with which the relation between the width of the feet and the downsizing would resemble a linear relation.

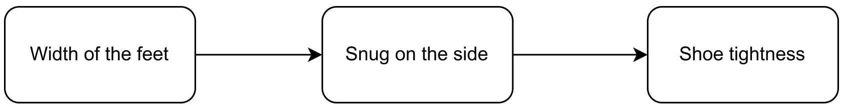

Figure 5.6: Diagram representing a possible relationship between the width of the feet and the downsizing.

It behaves differently analyzing the gender.

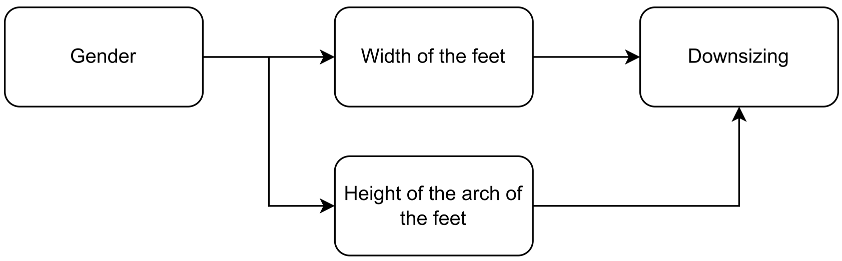

Figure 5.7: Diagram representing a possible relationship between the gender and the downsizing.

Differently from the males, females tend to have a narrower feet with a lower height of the arch of the feet. Those two relations are described from a linear relation. Though the downsize is inverse exponentially depending on the two previous parameters, meaning that:

$$
\begin{aligned}
\gamma &= c_1 \cdot x + c_2 \\
\rho &= c_3 \cdot x + c_4
\end{aligned}
\tag{5.8}
$$

where x is the gender variable. The relation with the downsizing could be described as:

$$
s = c_1 e^{\alpha \cdot x} + b e^{\beta x} + c = \theta \cdot e^{\phi \cdot x} + d \tag{5.9}
$$

Similarly it has been possible to classify the relation between the comfort level and the downsize.

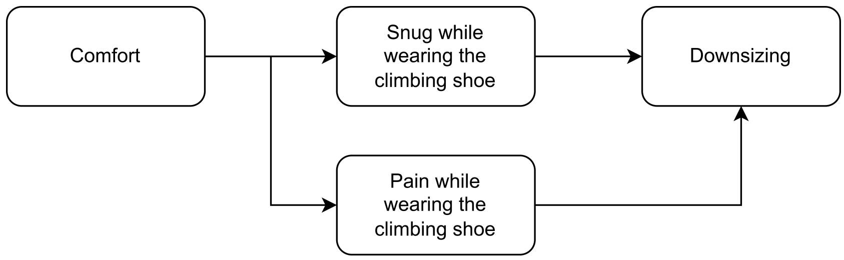

Figure 5.8: Diagram representing a possible relationship between the comfort and the downsizing.

where the relation between the comfort and the downsizing can be described from an inverse exponential while is linear with the snug while wearing the climbing shoe:

$$
\begin{aligned}
y &= c_1 \cdot e^{c_2 x} + c_3 \\
\gamma &= c_4 \cdot x + c_5
\end{aligned}
\tag{5.10}
$$

Finally the downsize relates to the pain and the snug with an exponentially inverse relation:

$$
\begin{aligned}
\rho &= c_6 \cdot e^{c_7 \cdot (c_1 \cdot e^{c_2 \cdot x})} + c_8 \\
\rho &= c_8 \cdot d^{c_7 \cdot x + c_5} + c_9
\end{aligned}
\tag{5.11}
$$

Those curves had been tuned testing and observing the outputs of each curve passed to the curve fit function.

5.2.1 Results

To verify the results of the approach through curve fitting, it was necessary to further elaborate the results of the dataset used in the study phase.
In fact it is necessary to use part of the data to carry out a training phase while a second part to test the predictive capacity of the dataset. The choice of how to split the data was made through a random choice and assigning a value of 1 to 20% of the dataset on a new column called 'training', while the remaining 80% was assigned the value 0. In this way, at each iteration the training dataset would remain unchanged. Furthermore, the data that are destined for the prediction phase would not have created dependencies in the training phase.
Having said this, it was possible to estimate the parameters of the curve estimated in the previous chapter using those that are the data obtained from the set of so-called training data. Subsequently it was possible to proceed with the prediction in the set of data provided.
At this point it was therefore possible to calculate the error of this model in the prediction phase with respect to the real values, hidden from the prediction model.
Accuracy relates to two variants: the first taking into account a half-size prediction error; the second instead considering 1 size of error. The reason for this choice lies on the fact that it has been estimated that a half size error can however be considered as acceptable by the interviewees. It was in fact pointed out that a half size less or more is not so penalizing as to justify a return at the store. As regards the full size, however, it was introduced for a purely theoretical reason and to verify the trend of the correctness of the prediction.

The accuracy reported for La Sportiva brand is:

- **50%** accuracy with a half-size margin
- **73%** accuracy with a 1 size error

while for La Sportiva aggressive models it has been computed:

- **48%** accuracy with a half-size margin
- **73%** accuracy with a 1 size error

Finally for La Sportiva Solution climbing shoe it has been computed:

- **42%** accuracy with a half-size margin
- **67%** accuracy with a 1 size error

5.3 Decision tree regressive model

5.3.1 Decision tree learning

A Machine Learning classifier is a type of algorithm able to assign a class label to a data input. This type of technology is a subset of Artificial Intelligence (AI), and has the ability to learn and improve its performance. Decision tree learning is one of the possible technologies that are part of machine learning.
An example for a machine learning approach, would be a classifier, an algorithm, that can distinguish a triangle from a square given a set of these objects. A classifier is able to associate the triangle images to the triangle class and squares images to the square class.
Generally speaking, a model of a classifier receives as input a set of data: from one part of the images labeled with the specific class they belong to; on the other hand, a portion of the data sent will be deprived of the labels and will be sent as input to the model in this way, to verify the goodness of the algorithm. The classifier is indeed able to assign the new unlabeled items received into the correct category.

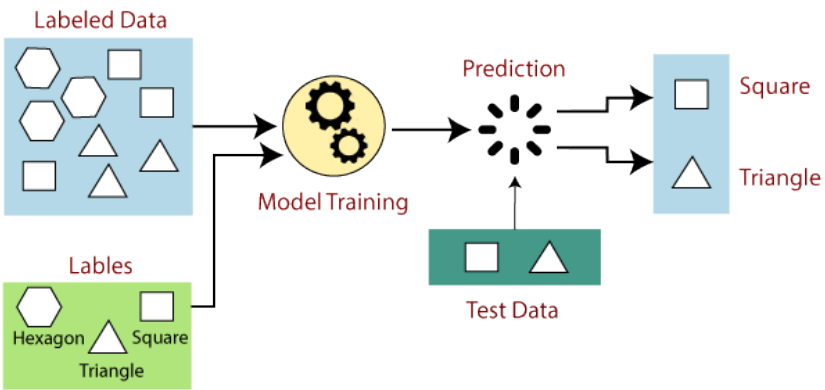

Figure 5.9: A machine learning classifier.

Each class of the classifier in our case could be represented as a downsize value. The vector $\mathbf{x} = x_1, x_2, x_3 \cdots$ defines the set of features that are used for the task of the decision tree. A classifier then, would watch the set of features from the new climber, and should be able to associate it to a specific class. The class would determine the suggested value that a new climber would fit with.

The decision tree learning we've been talking about so far, is a famous approach to describe a classifier but not only. It has been used in different areas, such as Data Mining or statistic. In our case it is possible to exploit a decision tree in the field of Machine Learning. It is considered a white-box predictor since the decision tree moves along one of two branches based on decisions that can be understood and analyzed. Visually, a decision tree model can indeed be represented as a rooted tree.[16]

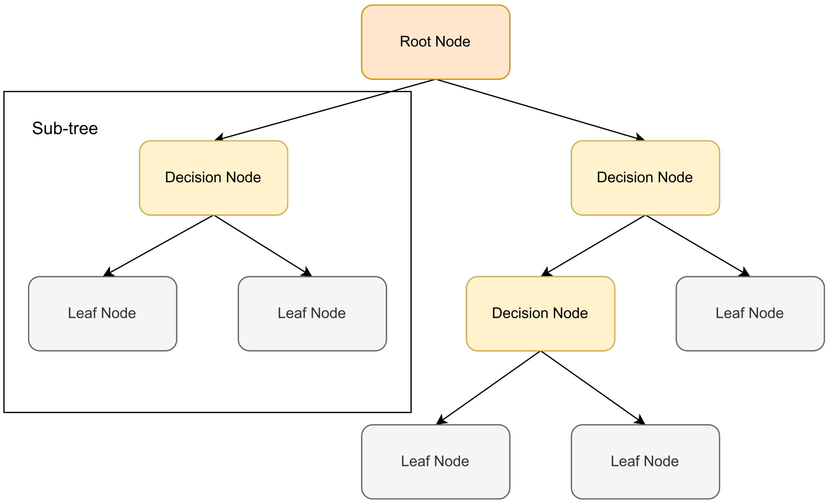

Figure 5.10: Decision Tree.

In a decision tree there are four different components:

1. **The Root Node**: it is the first node from which the graph starts to put out edges and leaves. The first node is associated with the first function that allows the algorithm to create a subdivision of space. It is therefore possible to return the suggested value from this first iteration.

2. **Decision nodes**: each node that is inside the tree, is called Decision Node. It refers to a node that allows to split the space of the results into two sub-spaces, according to a certain discrete function. The division is based on the attribute associated to the parent node.

[16]**Tree:** it is an abstract data representation that represents a set of hierarchically connected nodes. Each node can be connected to some children, while each child can each just one parent. If a node has not the parent, it is defined as root node.

3. **Leaves node**: all the nodes with no children[17].

Resuming the previous example, suppose we are building a decision tree to understand if the item passed is a triangle or a square. Assume that the parent node attribute refers to the number of sides:

- If the number of sides is equal to tree, then the class associated to the item would be classified as a 'triangle'
- If the number of sides is equal to four, the class associated would be classified as a 'square'

This type of decision made by the decision tree regressor is easily understandable. This allowed us to control and check the procedure with which the algorithm chooses the output. The partitioning takes into consideration one single attribute per time: if it is necessary to split a numeric value, for example, it is possible to considerate the splitting condition refers to a range of values.

To each leaf finally it is associated a value. This value is class that coincide with a target value. Alternatively it could be a probability with which the target having a certain value. The structure of a decision tree can be considered more or less complex given different metrics: the numbers of leaves, the number of nodes, the number of attributes or the depth of the tree. The decision tree module already implements a graph traversal algorithm capable of providing as output the optimal decision for each set of inputs.

5.3.2 Adopted model

With the intention to exploit the decision tree regressor, it has been decided to implement it. The predictor is taken from the library provided from Scikit Learn[18] under the supervised learning[19] section. Here it is possible to find a decision tree paragraph that relates with this approach.

It was decided to exploit just partially the incoming data. The questionnaire provided a lot of information that have been neglected both because they are not so relevant, but also

[17] **Node's children**: each relation between two nodes in a tree data structure is composed from a parent and a child. Each node in a tree has zero or more child, which are below it in the tree. A node that has at least a child is called parent node.

[18] **Scikit Learn:** it is an open source library that provide a set of efficient tools for the predictive analysis.

[19] **Supervised learning:** it is a subcategory of machine learning. It is characterised by a labeled dataset that allow the predictor to be trained. From that the model is able to learn and to predict new items labels.

because they did not seem to bring information to the ultimate goal of the prediction.

The data used had previously been encoded via one-hot encoding. However, it was not mandatory to use encoded data in this case, since the decision tree regressor is able to encode categorical variables as input variables. Other date had been considering while testing the predictive model, but no improvements were tracked. The considered features are:

- the expertise
- the frequency
- the grade
- the foot shape
- the toe shape
- number of normal shoe size
- downsize value

To implement a machine learning algorithm it is fundamental to split the dataset.
In machine learning it is vital the study and construction of algorithms that can learn from and make predictions on new data. To accomplish this task, it is necessary to preserve a subset of the available data as unseen from the predictor. For such a reason it has been divided the data into those two sets.
From the training set it is possible to extract another subset, called Validation set. The validation set is taken from a portion of the training set, to provide a comparison meter to the training set. So to have real predictive ability, we give our model data that it has never seen and we tell it to make a prediction. This data must necessarily be labelled to be compared. It is called validation set because it deals with validating the results obtained in the training set. If the performances were scarce, we would have to go to modify the model that is being used.

The test set, on the other side, does not need any label since it should be a set of data to test and visualize the performances of the model.
The training set is passed together with the downsize measure to the decision tree regressor. This process, called training, would create a model that learns to elaborate the inputted

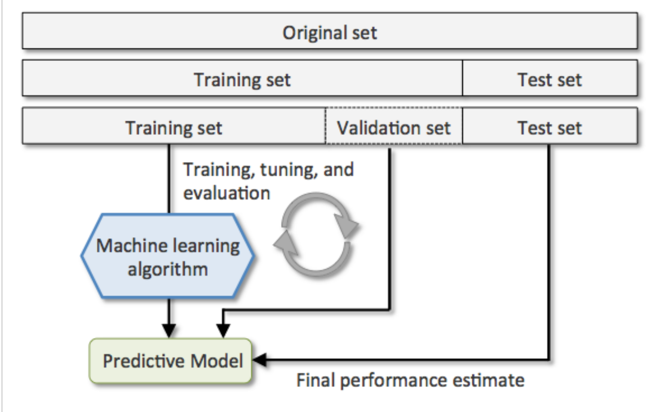

Figure 5.11: Division into Training, Validation and Test sets.

data, and to suggest the best outcome to copy the already seen data. The model is then able to associate a value to each different input for the validation set, based on the most probable downsize measure for each climber. It is then possible to compute the grade of the accuracy based on the actual downsize value.

Before calculating the effective accuracy of the model, it was decided to increase its strength by improving the parameters passed to the predictive model. Furthermore, to have a greater number of values on which to train our model, the practice of cross validation on the input dataset was adopted. Beyond that it was necessary to verify the goodness of the decision tree regressor: to do this it was necessary to perform a tuning of the hyperparameters.

Cross Validation The first practice applied before providing the data to the predictor is called Cross Validation.

This method would use different portions of data to train or validate the model during each iteration. As said, the training set can't coincide with the validation set, since the model should be able to associate a class to a new input without ever having seen it before. Though if the set of training data are too few, it could be hard to have a well performing model.

Along each iteration the set of data are divided with on a given percentage on a validation fold, while the remaining falls into the training. The function comes from the Scikit Learn library, the model selection library. The number of splits indicated as parameter of the function, coincide with the percentage of data that are assigned to the validation set and to the training set. The default values suggested is $n = 5$ splits: it mean that 20% of the training data would be assigned to the validation while, the 80% are used to train the model.

This methods is both used to avoid an overfitting issue or distortion due to the selected

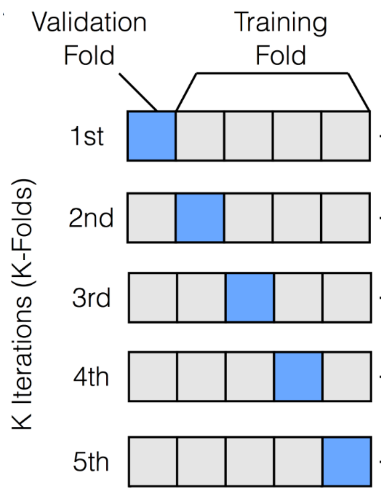

Figure 5.12: K-Fold cross validation.

set of training. The overfitting issue comes into place when trained model is too much conditioned from the seen data and is not able to judge new data and fails predictions.

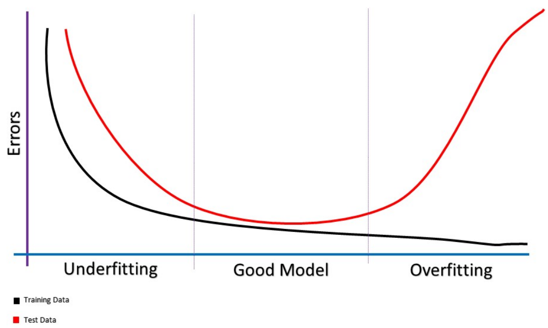

Figure 5.13: Overfitting graph.

In the figure 5.13 it is possible to see that along the y axis it is placed the error that the model accomplish during the prediction phase. Along the x axis is instead present the goodness of the predictive model. As written in the legend, the error while predicting training data would rather diminish while in the overfitting zone. The model would be hardly conditioned from the seen data and it is not able to generalize the seen results. On the other side the error in the test set, would drastically increase. To avoid this situation is though applied the K-Fold approach that increase the data adopted for training, thanks to those $k-1$ fold used n times to train the model.

The distortion issue comes into place if the selected training items would randomly describe an unwanted specific behaviour of the data used for the training. The K-fold would explore all the training data.

Hyperparameter tuning The second practice adopted is the tuning of the hyperparameters.
The model provided from the library of Scikit Learn presents different parameters that could be set based on the necessities.
Some parameters are, for example, the *criterion* and it is used to estimate the quality of the split. The *splitter*, that's used to select the strategy to split each node. The *max_depth* of the decision tree used to accomplish the prediction and so on.
Changing all those parameters, the model would be able to better learn and predict the data from a new given set. Nevertheless, it would be impossible to test all the possible values manually, it is necessary to adopt an automatic approach. It has been decided to exploit the GridSearch approach to search for the best parameters from a given set.
GridSearchCV is a module provided from Scikit Learn library. It provide a container that accept the model used for the prediction together with a set of different parameters that we are interested into. For example, it would be interesting to exploit the available parameters for the *criterion*, it would be possible to pass to the gridsearch a set of possible values such as "squared_error", "friedman_mse" or "absolute_value" and the container would take care to test all the possible combinations to find out the best solution to build the model.

This solution would increase the combinations dramatically if the number of inputted parameters are too high, since the module needs to test all the possible available combinations. On the other side, it would spare a lot of manual work. Once the library has explored all the possible combinations and parameters, it will return the solution that perform the best.

The model was encapsulated into the GridSearchCV framework together with a variety of different parameters. The model was then tested on a test set. It has been returned the accuracy of the model.

5.3.3 Results

The trained model ran over the validation set and predicted the downsize measure to each climber, based on the information retrieved from the questionnaire.
To estimate the goodness of the model, it had been looked for the true values returned from the questionnaires and the output of the prediction. It was counted the accuracy based on the distance with respect to the right prediction: if it was correct within an half-size error,

within a one full size error, within one and half size error and so on.

It would considered a relevant result if the model would be able to estimate the size within a half-size error. The results are reported in the figure 5.14.

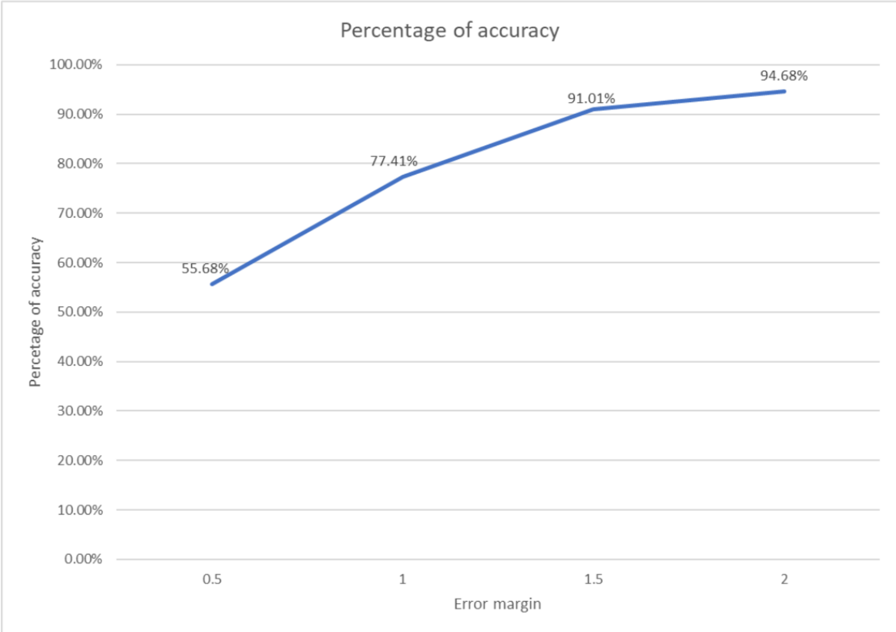

Figure 5.14: Accuracy results.

The model is able to predict the right size of the climber with an accuracy of half-size with the **55.85%** of the probability. It increases to **77.41%** with a 1 full size error. The accuracy would increase more and more looking for a bigger margin.

Chapter 6

Next developments

6.1 Computer Vision injection

This e-commerce project was initially studied and developed with the intention to create a predictor able to provide to a customer the best size of climbing shoe to limit the percentage of returns.
The results obtained with the questionnaire approach left some improvement room. The predictive model built and described in the previous chapter, defined a baseline to start from. On the other side, a 7 minutes long questionnaire is not something that would fit very well in an e-commerce that is intended to provide a very fast and enjoyable service. It is then necessary to retrieve some of those information more quickly.

More than that, some others issues come into place. The sneaker shoe size is something that is not reliable as a reference measure. While the feet size is something that can be computed in centimeters/inches, the number of the sneaker is indeed something that vary based on the preferences of the user. Also the size would depend on the model of sneaker that a person is using. It is commonly known that two different brands do not accommodate the same feet with the size shoe size. Someone wearing 43 of a Nike shoe, would probably fit well in a 44 of another brand, and so on. The shoe size asked to the user while taking the questionnaire could be substituted with the actual measure of the feet length. Beyond this, it would be possible to identify the type of toe by classifying the shape of the profile of the toes into the three classes used to identify them [Figure 3.9].

The most plausible approach would exploit the Computer Vision field.
It would be possible to inject a Computer Vision engine able to extract the feet length and width starting from the image of the feet. Simultaneously it possible to retrieve the toe

shape category just taking a look on an uploaded image. This procedure could be accomplished starting from a standard reference measure, such as an $A4$ paper placed under the foot, and rectifying the image with respect to the paper. It is possible to use it as a reference and compute the length based on the length of the paper, that is known to be $29.7x21$cm. Referred to the toe shape, given a white background due to the paper, it would be possible to split the feet from the background thanks to an edge detector [20]. Since each type of toe shape can be recognised from the silhouette, it is possible to create a set of binary images for each of the classes available: Egyptian, Roman and Greek.
At that point it would be a matter to classify the input image into one of those 3 categories, to retrieve the information used in the previous model.

This approach underlines different advantages on user and technical sides. The user is no more asked to answer to questions based on their opinions. The information would became objective and reliable for any next improvement of the actual predictive model. The user is asked to take a picture from the phone, that is something that is very straightforward and immediate instead of answering questions through a questionnaire. The information gleaned from the images can improve the efficiency and a further approach can be applied on different measures having a longer time perspective. By collecting images from different angles it may be possible to extract other information, such as, for example, the height of the instep.

It is indeed necessary to guarantee to the customers that the images gathered and processed are safe and that the company will respect an ethical treatment of the collected images.

Before the first release, these practices were implemented to support the previous algorithm.
Alongside the predictive model, it had been implemented the possibility for the user to upload the images to avoid some questions about their feet sizes. The upload of the images with an A4 or Us Letter paper under the feet would allow the algorithm to understand the length of the feet and to convert it into a size of sneaker shoes. Together with this information it had been gathered also the width of the feet and the shape of toes through a classifier.

[20]**Edge detector:** it is a set of mathematical methods that could work on a image to identify edges

6.2 ANN: Artificial Neural Network

The predictive model adopted in the previous chapter obtained a good result. On the other side it showed some limits on the capability to understand and predict the right value.
The decision tree regressor is not indeed a very powerful predictor, and lack on the capability to capture some undergoing relations that could exists. A possible approach to be exploited would lie on a black box approach. In that case it could be possible to exploit, for example, an Artificial Neural Network.
A Neural Network (NN), or Artificial Neural Network, is a computer system inspired by the biological functioning of the animal brain. It is composed by some components, called neurons, that try to resemble a real neuron. It create some connections and it transmits information to other neurons.

A neural network is made of neurons. Neurons are organized into layers. Each neuron of each layer could be connected to a neuron of the next layer. The first layer is called input layer and it receives the inputs. The last one is called output layer and it produce the output of the neural network. The way the neurons transmit information is related to a non-linear function: a neuron receives an input signal. It is asked to elaborate it. If the function returns that a neuron has to fire, it will fire the output value. Also the learning process of a neural network deals with a *weight* measure: if the weight decrease, the strength of the connection between two neurons decrease; on the other side if the weight increase the connection would be tougher.
A neural network structure is extremely flexible and allow to exploit different task efficiently. The layers organization would allow a huge amount of neurons, that could be fully connected, partially connected or sometimes, to avoid overfitting, some of them are dropped from the network randomly. Each neuron depict a new parameter that would and it would describe a network able to solve more complex task.

Avoiding going into too many details, there are few advantages of using a Neural Network that are worth the mentioning.

- Neural networks are built to be constantly learning. Once a neural network is trained, it can trained more to be more performing and effective.
- It is able to understand an underneath relationship that cannot be directly caught on a preliminary phase.

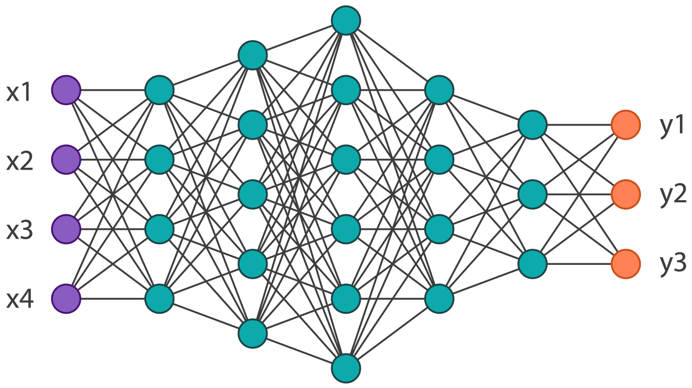

Figure 6.1: A Neural Network.

- A Neural Network would work similarly to a very small human brain, but would outperform it based on the time that it could work.
- A Neural Network would complete task quickly, which can take humans longer time.

It could be interesting to explore this topic further, to understand if an approach like this is able to provide a boost on the accuracy on the model previously adopted. Assuming that a Neural Network would be able to understand some undergoing relations, it is imaginable that it will be able to overcome the results obtained with the decision tree regressor.

Chapter 7

Conclusions

Since the beginning, the project had set out to advance conclusions about the possibility of creating a piece of technology that could create added value both from a technical point of view and from a user experience point of view.
And in fact, the goal of providing the user with an innovative experience in the act of buying a new pair of shoes has been achieved, creating a new standard for what concerns the choice of size online: the user is not more called to interpret a set of rules and instructions to identify the most suitable size, but he is asked to accept a suggestion based on his own physicality and the buyer's history.
On the other side it is intended to be pragmatic. It has to push further the capability of user to buy with more awareness a product. Under this assumption the user should record fewer returns. On that side it has not been possible to test if the e-commerce structure would help. On the positive side, it is possible to conclude that the results obtained seems optimistic.

From the point of view of user experience, not yet enough interactions between users and the new online purchasing method have been collected to be able to conclude that it was positive.
On the other hand, there was positive feedback on the small sample of users who were able to experience an alpha version of the entire e-commerce pipeline. In general, some users have positively reported that the possibility of uploading photos and leaving the identification and association of the width of the foot and the tip of the foot to the algorithm has solved a significant problem. Some users have indeed reported their difficulty in falling back into one of the classes listed. Just as some users have pointed out that reporting the shoe size worn is usually not always that simple, as for different shoe brands, the size used

can even vary by a whole size.

As far as the technical part and the development of the predictive model are concerned, the results were interesting, but still limited.
The accuracy percentage of the model is not satisfactory to imagine having obtained a definitive result. On the other hand, this preliminary study has highlighted the boost it could have with a more focused and long-term investigation. It is important to take into consideration that, in this analysis of the case, climbing shoes have been treated, a particularly delicate sector and with great subjectivity. It is conceivable that an approach to a less complex sector steeped in personal opinions would have led to more promising results.

Acronym

UX User Experience

UI User Interface

CSV Comma-Separated Values

CF Collaborative Filtering

GMM Gaussian Mixture Model

TF-IDF Term frequency–inverse document frequency

NN Neural Network

Bibliography

[1] *Anatomy of a Climbing Shoe.* https://en.paperblog.com/anatomy-of-a-climbing-shoe-1865306/.

[2] Sheena Butler-Young. *Why $62 billion In Apparel Footwear Purchases Are Returned By Consumers.* https://footwearnews.com/2016/business/retail/why-consumers-return-shoes-clothes-purchases-footwear-apparel-retail-online-226642/. Jun 3, 2016.

[3] *How Tight Should Climbing Shoes Be?* https://www.climbingshoereview.com/how-tight-should-climbing-shoes-be/.

[4] *La Sportiva brand.* https://www.lasportiva.com/en. Jun 3, 2019.

[5] Haiko Schulz Michael Meyer Sascha Dietrich and Anke Mondschein. *Comparison of the Technical Performance of Leather, Artificial Leather, and Trendy Alternatives.* Feb 13, 2021.

[6] Lukas Peters. *nike.com: E-Commerce net sales from 2014 to 2022.* https://www.statista.com/forecasts/1218320/nike-revenue-development-ecommercedb. Apr 30, 2022.

[7] D. Tighe. *Number of Nike stores globally 2009-2021.* https://www.statista.com/statistics/250287/total-number-of-nike-retail-stores-worldwide/. Jul 27, 2022.

[8] *Types of Climbing Shoes.* https://www.climbingshoereview.com/types-of-climbing-shoes/.

[9] Thomas Küpper Volker Schöffl. *Feet injuries in rock climbers.* Oct 18, 2018.

[10] *Words vectorization.* https://towardsdatascience.com/art-of-vector-representation-of-words-5e85c59fee5#:~:text=Words. Jun 3, 2019.

[11] Anna Zinola. *Reso gratuito: da Zara a Uniqlo, è finita l'era in cui si restituiva l'acquisto senza pagare.* https://www.corriere.it/economia/moda-business/cards/reso-gratuito-zara-uniqlo-finita-l-era-cui-si-restituiva-l-acquisto-senza-pagare/commissioni_principale.shtml. Aug 28, 2022.

Ci tengo a ringraziare la Prof.ssa Federica Battisti
per avermi accompagnato durante questa
tappa del mio percorso universitario.

Ringrazio la mia famiglia e Valeria,
elementi imprescindibili di questo traguardo.

Ringrazio Pier Mattia, Daniel, Giulia, Valentina, Lorenzo, Massi
e tutto il team di Uqido,
per avermi seguito e permesso di costruire
quest'ultimo tassello assieme.