

▶▶
UHASSELT



Maastricht University

KNOWLEDGE IN ACTION

Faculteit Wetenschappen **School voor Informatietechnologie**

master in de informatica

Masterthesis

Interactieve exploratie van herstelstatus van recreatieve atleten op basis van machine learning en smartwatch data

Steve Timmermans

Scriptie ingediend tot het behalen van de graad van master in de informatica

PROMOTOR :

Prof. dr. Gustavo Alberto ROVELO RUIZ

COPROMOTOR :

dr. Eva GEURTS

BEGELEIDER :

Mevrouw Anouk MICHIELS

De transnationale Universiteit Limburg is een uniek samenwerkingsverband van twee universiteiten in twee landen: de Universiteit Hasselt en Maastricht University.



UHASSELT

KNOWLEDGE IN ACTION

www.uhasselt.be

Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

2023
2024



Maastricht University

Faculteit Wetenschappen

School voor Informatietechnologie

master in de informatica

Masterthesis

Interactieve exploratie van herstelstatus van recreatieve atleten op basis van machine learning en smartwatch data

Steve Timmermans

Scriptie ingediend tot het behalen van de graad van master in de informatica

PROMOTOR :

Prof. dr. Gustavo Alberto ROVELO RUIZ

COPROMOTOR :

dr. Eva GEURTS

BEGELEIDER :

Mevrouw Anouk MICHIELS

Dankwoord

Gedurende een compleet jaar heb ik gewerkt aan het afronden van deze masterproef. Dit heb ik zeker niet alleen gedaan, daarom wil ik de volgende bijzondere mensen en organisaties bedanken voor de ondersteuning bij mijn werk.

Allereerst wil ik de onderwijsinstelling UHasselt bedanken, waar ik deze opleiding heb mogen volgen. Mijn speciale dank gaat uit naar mijn promotor, Prof. Dr. Rovelo Ruiz Gustavo, mijn co-promotor, Dr. Eva Geurts, en mijn begeleider, Mevr. Anouk Michiels, voor hun onmisbare adviezen gedurende mijn hele traject. Daarnaast wil ik hen bedanken voor het beantwoorden van al mijn vragen en het geven van feedback tijdens het schrijven van mijn thesis.

Een bijzondere dank gaat uit naar mijn goede vriend en oud-collega, Dhr. Sean Snellinx, voor het lenen van zijn ontworpen tool om de dataverwerking te faciliteren. Zijn hulp was essentieel voor het voltooien van deze masterproef.

Ik ben ook mijn vrienden dankbaar voor het testen van mijn applicatie en het wijzen op concepten die onduidelijk waren en verbeterd moesten worden. Mijn respondenten voor de enquêtes wil ik bedanken voor hun tijd en moeite. Zonder hen was het niet mogelijk geweest om tastbare resultaten neer te zetten.

Tot slot wil ik graag mijn mama en tante bedanken voor hun steun en verwennerijtjes tijdens het afgelopen academiejaar.

Bedankt allen!

Samenvatting

Deze masterthesis onderzoekt het herstel van recreatieve atleten door het gebruik van machine learning en data van een smartwatch. Het doel van het onderzoek is om het herstel van recreatieve atleten nauwkeurig in te schatten, zodat zij optimaal kunnen trainen. Daarnaast wordt beoogd om atleten tips en meer inzicht te geven in de componenten van herstel, waardoor hun trainingsprestaties kunnen verbeteren. De datacollectie is uitgevoerd met Fitbit smartwatches, waarbij verschillende parameters zoals hartslag in rust, hartslagvariabiliteit, en slaapgegevens zijn verzameld. Machine learning-classificatie werd vervolgens toegepast om te bepalen wat de herstelstatus was van de atleet, één dag na zijn training.

De resultaten van de *user study* toonden aan dat de ontwikkelde webapplicatie functioneel was, hoewel er ruimte voor verbetering in het ontwerp en gebruiksvriendelijkheid bleef. De machine learning-modellen gaven een hoge nauwkeurigheid, maar andere prestatieparameters suggereerden dat er nog aanzienlijke verbeteringen nodig zijn.

Uit de conclusies blijkt dat verder onderzoek noodzakelijk is om de webapplicatie te optimaliseren, zodat sporters de herstelcomponent niet negeren in hun training. De machine learning-resultaten tonen aan dat het inschatten van het herstel van een atleet mogelijk is, maar dat de huidige implementatie nog verfijnd moet worden voor praktische toepassingen. Bovendien wordt aanbevolen om de dataverzameling uit te breiden naar een groter aantal proefpersonen en meer datapunten te verzamelen, om de robuustheid en generaliseerbaarheid van de resultaten te vergroten. Hierbij zou ook de integratie met diverse smartwatch-modellen moeten worden overwogen, zodat elke sporter, ongeacht het type smartwatch dat hij gebruikt, kan profiteren van de herstelaanbevelingen en inzichten van de webapplicatie.

Tegelijkertijd moet het design van de webapplicatie verder worden verbeterd om een aantrekkelijke en intuïtieve gebruikerservaring te bieden, waardoor sporters gemotiveerd worden om de applicatie regelmatig te gebruiken. Dit stelt hen in staat om hun training verder te optimaliseren.

Inhoudsopgave

1	Inleiding	6
2	Optimalisaties voor prestaties in wielrennen	8
2.1	Optimale verdeling trainingsintensiteit	8
2.2	Stretchen	10
2.3	Krachttraining	11
2.4	Energiesystemen voor een duursporter	12
2.4.1	Koolhydraatoptimalisaties	12
2.5	Supplementen	13
2.6	Herstel	16
2.7	Bestaande applicaties	18
2.7.1	Garmin horloges	18
2.7.2	Whoop	19
2.8	Bepaling van fysiologische stress van een sportactiviteit	20
3	Machine learning	21
3.1	Wat is machine learning	21
3.1.1	Onderverdelingen	21
3.1.2	Tree modellen	22
3.1.3	Boosting modellen	23
3.2	Machine learning in de praktijk	24
3.3	Verschillende evaluatie metrieken	25
3.4	Kruisvalidatie	27
4	Verzameling gegevens voor inschatting herstelstatus	28
4.1	Datacollectie	28
4.1.1	Meetbare parameters	28
4.1.2	Procedure van de datacollectie	30
4.2	Verwerken van de data	31
4.3	Trainingsbelasting berekenen en verbeteren	32
4.4	Analyseren, opschonen en verbeteren van de dataset	35
4.5	Aanpassen dataset aan gekozen machine learning	36
4.6	Test procedure voor machine learning resultaten	38
5	Resultaten	39
5.1	Resultaten zonder optimalisaties	39
5.2	Geoptimaliseerde <i>hyperparameters</i>	40
5.3	SMOTE en geoptimaliseerde <i>hyperparameters</i>	43
6	Webapplicatie: voorspellen van herstelstatus recreatieve atleet	46
6.1	Verkennen van de dataset en eenvoudige relaties	46
6.2	Verkennen van complexe relaties	47
6.3	Introductie van scenario's	48

<i>INHOUDSOPGAVE</i>	5
6.4 Inputwaarden voor voorspelling	49
6.5 Voorspelde herstelstatus en tips	49
7 User Study	51
7.1 Testprocedure	51
7.2 Resultaten enquêtes	52
7.2.1 Algemene enquête	52
7.2.2 System Usability Scale (SUS)	54
7.2.3 User Experience Questionnaire (UEQ)	55
8 Future Work	58
9 Conclusies	59
9.1 Antwoord op research questions	59
9.2 Zelfreflectie	60

Hoofdstuk 1

Inleiding

De markt voor sportprestatie-analyse om sportieve prestaties te kwantificeren, groeit gestaag. Deze trend, in combinatie met de opkomst van draagbare technologie, biedt zowel topatleten als amateuratleten de mogelijkheid om tijdens en na het sporten kwaliteitsvolle en nuttige informatie te verzamelen [1]. Het succes van apps zoals Strava¹ illustreert dat atleten deze kans benutten. In 2016 had Strava slechts 20 miljoen gebruikers, maar in 2023 rapporteerde het platform al 120 miljoen gebruikers, een verzesvoudiging in slechts zeven jaar tijd [2]. Strava is echter niet het grootste platform; concurrerende platforms zoals Google Fit² en Apple Health³ meldden elk al 350 miljoen gebruikers [2].

Een probleem voor recreatieve wielrenners is dat al deze platformen niet altijd een goede integratie bieden om een totaalbeeld van de training of de gezondheidsstatus van de atleet te verkrijgen. Algemene apps zoals Google Fit app of de Fitbit⁴ app, die vaak met een smartwatch wordt geleverd, bieden vaak slechts een generiek overzicht voor sporters, beperkt tot gezondheidsgegevens of enkele basisstatistieken van hun sportactiviteiten. Het is bijvoorbeeld niet altijd mogelijk om sensoren, zoals een vermogensmeter, aan deze applicaties te koppelen om verder inzicht te krijgen in de geleverde training. Daarnaast missen deze apps vaak de noodzakelijke tools om trainingen uitgebreid te analyseren, zoals platformen die meer gericht zijn op het fietsspecifieke aspect. Fietsplatform zoals Strava bieden de mogelijkheid om uitgebreide trainingsprogramma's op te stellen en hun trainingen optimaal te plannen. Na elke activiteit worden diverse parameters en grafieken gepresenteerd, waardoor de atleet een diepgaand inzicht krijgt in zijn of haar trainingen en de effectiviteit ervan kan beoordelen. Bovendien bewaken deze geavanceerde platformen de stress en vermoeidheid van atleten op basis van de vastgelegde activiteiten, om te voorkomen dat ze zich overbelasten. In het geval van wearables wordt de inspanning doorgaans geschat aan de hand van parameters zoals de hartslag of de snelheid tijdens hardloop- of fietsactiviteiten [3]. Een dergelijke schatting biedt een benadering van de fysiologische stress die het lichaam kan ervaren als gevolg van sportieve activiteit. Bijgevolg is het cruciaal dat het lichaam de mogelijkheid heeft om te herstellen en zich voor te bereiden op een nieuwe trainingssessie. Bij het beoordelen van het herstelproces schieten de platformen zoals Strava tekort. Deze platformen maken gebruik van wetenschappelijk gevalideerde formules om de stressbelasting van de training op het lichaam te schatten. Het is echter essentieel op te merken dat het menselijk lichaam geen machine is, waarbij een specifieke input altijd een voorspelbare output oplevert. Bovendien bestaat er aanzienlijke variabiliteit in de fysieke conditie van gebruikers van deze platformen, variërend van professionele atleten tot recreatieve sporters.

¹<https://www.strava.com/>

²<https://www.google.com/fit/>

³<https://www.apple.com/ios/health/>

⁴<https://www.fitbit.com/global/be/home>

Wat in het algemeen opvalt bij de meeste applicaties is dat er vaak te weinig aandacht wordt besteed aan het herstelaspect na een training. Sommige applicaties kunnen wel een inschatting geven van de hersteltijd die het lichaam nodig heeft. Dit is bijvoorbeeld een kenmerk van Garmin-horloges⁵. Deze horloges passen de geschatte hersteltijd aan op basis van de slaapkwaliteit; bij een slechte nachtrust wordt de hersteltijd verlengd en bij een goede nachtrust verkort. Dit is gebaseerd op actuele wetenschappelijke literatuur waaruit blijkt dat slaap een onmisbare factor is in het herstelproces van het lichaam na intensieve fysieke inspanning. Een onvoldoende hersteld lichaam is namelijk niet in staat om optimaal te presteren tijdens de volgende trainingssessies. Gebrekkige slaapkwaliteit kan verschillende nadelige effecten veroorzaken, zoals het vertragen van het herstel en de reparatie van spierweefsel, verminderde cognitieve prestaties en veranderingen in de perceptie van pijn [4].

Hoewel de literatuur duidelijk beschrijft hoe slaap een sterke invloed heeft op het herstel van een sporter, blijven de meeste smartwatches vaag over de methoden waarmee het herstel op hun platform wordt ingeschat. Door minder transparant te zijn, wordt het voor atleten moeilijker om hun herstel te verbeteren, omdat ze niet concreet weten welke maatregelen ze moeten nemen. Het zou beter zijn als deze platforms individuele tips zouden geven aan elke atleet, afgestemd op hun specifieke situatie en behoeften.

Uit bovenvermelde tekortkomingen in de huidige technologie, ontstaan de volgende onderzoeksvragen die deze thesis tracht op te lossen:

- In hoeverre kan een wearable een accurate schatting maken omtrent fysiologische stress van sportactiviteiten van een atleet?
 - Kan er met behulp van data van een wearable ingeschat worden wanneer een atleet voldoende hersteld is voor een nieuwe training?
 - Is het mogelijk om voor elke individuele atleet een gepersonaliseerd herstelvermogen te bepalen?
- Welke methodes zijn het meest gebruiksvriendelijk om data omtrent trainingen en herstel te visualiseren?

⁵<https://www.garmin.com/en-US/>

Hoofdstuk 2

Optimalisaties voor prestaties in wielrennen

Het professionele wielrennen is voortdurend bezig met innovatie en verdere optimalisatie van de sport, zoals het verbeteren van trainingsstrategieën. Na verloop van tijd zullen deze nieuwe optimalisaties geleidelijk ook hun weg vinden naar recreatieve wielrenners. Hoewel optimalisaties voor recreatieve sporters niet dezelfde urgentie hebben als voor professionals, die wedstrijden moeten winnen voor hun inkomen, kunnen alle recreatieve sporters extra snelheid of uithoudingsvermogen waarderen, vooral als dit zonder extra inspanning te realiseren is. Dit is met name relevant omdat de meeste recreatieve sporters een vaste baan en diverse andere verplichtingen hebben. Hierdoor zijn zij beperkt in het aantal trainingsuren dat zij kunnen besteden. Optimalisaties kunnen zeker van pas komen wanneer ze specifieke doelen voor zichzelf hebben gesteld, zoals een grote of speciale (toer)tocht, al dan niet met andere sporters. Denk hierbij aan de Ventourist, de Cannibalette of de Cannibale¹, die worden georganiseerd als onderdeel van een bewegingscampagne van Sporta. Het stellen van dergelijke doelen fungeert als motivatie voor hun trainingsinspanningen. Om met succes zo'n veeleisende uitdaging aan te gaan, is het raadzaam om voldoende voorbereid te zijn. Het kan daarbij behulpzaam zijn om optimalisatiestrategieën van professionele wielrenners (gedeeltelijk) over te nemen.

2.1 Optimale verdeling trainingsintensiteit

Net zoals recreatieve sporters, zijn professionele wielrenners ook onderworpen aan een beperkt aantal uren dat zij ter beschikking hebben om te trainen. Algemeen genomen hebben professionele wielrenners meer tijd gereserveerd om trainingen af te werken, dit wilt immers niet zeggen dat zij willekeurig trainingen afwerken. Binnen de professionele sportwereld wordt consequent gewerkt met gedetailleerde plannings en trainingsblokken om een atleet optimaal voor te bereiden voor één of meerdere specifieke wedstrijdendagen. Om een effectieve communicatie tussen de coach en zijn atleet te faciliteren, is er behoefte aan een mechanisme waarmee de coach kan verduidelijken hoe intensief een training moet worden uitgevoerd. Hier komen modellen van trainingzones aan te pas. Deze modellen verdelen het spectrum van geen inspanning tot maximale inspanning in verschillende zones. Alle modellen bevatten niet noodzakelijk allemaal evenveel verschillende zones. Daarnaast baseren ze zich op verschillende soorten criteria om de grenzen van elke zone te bepalen. De criteria's kunnen zijn : gevoel, vermogen, lactaatniveau, hartslag. Voor alle criteria bestaan er verschillende testprotocollen of inschattingen om een relatief correcte benadering te maken. Professionele wielrenners maken vaak gebruik van lactaatniveaumetingen in een laboratoriumomgeving. Tegelijkertijd worden hier ook vermogen- en hartslagmetingen uitgevoerd om een zeer nauwkeurig profiel van de atleet op te stellen.

¹<https://www.sportamonventoux.be/>

Intussen kunnen recreatieve atleten deze metingen ook laten uitvoeren, maar deze kosten enkele honderden euro's ². De recreatieve sporter moet voor zichzelf bepalen of hij dit de moeite waard vindt. Als hij beschikt over een hartslag- of vermogensmeter, kan hij nog steeds een goede benadering van zijn trainingszones verkrijgen door thuis een inspanningstest te doen met een vermogensmeter, of door gebruik te maken van online tools waar hij zijn rusthartslag en maximale hartslag kan invoeren om zijn trainingzones te krijgen. Als de recreatieve sporter geen idee heeft van zijn maximale hartslag, kan een schatting op basis van zijn leeftijd worden gemaakt ³.

Een simpel maar bekend voorbeeld is het driezonesmodel van Seiler uit 2006 [5]. De grenzen van de zones binnen dit model worden bepaald door twee lactaatgrenzen LT1 en LT2, die empirisch vastgesteld zijn door wetenschappelijk onderzoek. LT1 is een maximale aerobe grens waar de lactaatproductie relatief laag en stabiel blijft, typisch rond de 2 mmol/l. LT2, de anaerobe lactaatgrens, wordt gekenmerkt door het kantelpunt waar de lactaatproductie niet langer lineair toeneemt, maar exponentieel. Dit punt werd vastgesteld rond de 4 mmol/l. concrete betekent dit voor een sporter dat activiteiten in zone 1 als rustig moeten worden beschouwd. De tweede zone komt overeen met een gematigde intensiteit en de derde zone is de zone waarin een zeer intensieve inspanning plaatsvindt.

Naast de modellen voor trainingzones zijn er classificaties om een hele trainingsperiode in categorieën in te delen op basis van de tijd die een atleet in elke zone gespendeerd heeft. Binnen deze masterproef zullen we ons enkel beperken tot het model van Seiler om te classificaties te berekenen. Met deze aanpak kunnen er vijf verschillende classificatie modellen onderscheiden worden: Base, Pyramidal, Threshold, Polarized en HIIT. Deze vijf opties krijgen een label toegekend aan de verdeling van de trainingsintensiteit. Naast deze opties kan de atleet ook een unieke belasting creëren, die vervolgens onder een unieke samenstelling wordt geïnclassificeerd.

Binnen de literatuur zijn er geen strikte grenzen voor elke classificatie te vinden. Echter een studie biedt een wiskundige formule die de mate van "polarized" training uitdrukt. In dit onderzoek worden tevens de volgende richtlijnen gepresenteerd, gebaseerd op een model met drie zones [6].

- Base: een Pyramidal uitzicht met stevige nadruk op Zone 1 (Bv: 85 - 10 - 5).
- Pyramidal: %Zone 1 > %Zone 2 > %Zone 3 (Bv: 70 - 20 - 10)
- Threshold: %Zone 1 > %Zone 2 > %Zone 3 of %Zone 2 > %Zone 1 > %Zone 3 Nadruk moet op Zone 2 liggen, maar kan ook een Pyramidal uitzicht hebben (Bv: 40 - 50 - 10 of 50 - 45 - 5).
- Polarized: %Zone 1 > %Zone 3 > %Zone 2 en met liefst zo weinig mogelijk Zone 2 (Bv: 80 - 5 - 15).
- HIIT: %Zone 3 > %Zone 1 ≥ %Zone 2. Nadruk vooral op Zone 3 (Bv: 20 - 10 - 70).

Een atleet heeft verschillende mogelijkheden bij het plannen van zijn trainingen. Het zou bijzonder interessant zijn als er een optimale verdeling van trainingsbelasting bestond zodat atleten op de meest efficiënte manier naar hun eigen doelen kunnen trainen. Zeker voor recreatieve atleten met beperkte tijd door verschillende andere dagdagelijkse verplichtingen kan dit relevant zijn. Uit een onderzoek in 2019 [7] blijkt dat een trainingsplan met het grootste deel van de belasting in zone 1, garant staat voor meeste succes in sportieve uitdagingen waar het uithoudingsvermogen cruciaal is. Hierdoor kunnen de classificaties Threshold en HIIT uitgesloten worden als optimale voorbereidingen. Twee andere onderzoeken [8], [9] tonen aan dat enkel een hoge intensiteit de VO2max grenzen en daarbij ook de prestaties van de atleet kan verbeteren. Hoewel deze strategie toch een 20-tal jaar geleden gebruikt werd door profwielrenners volgens een studie uit 2015 [10].

²<https://www.adlon.be/sportmedische-testen/inspanningstest>

³<https://www.myprocoach.net/calculators/hr-zones/>

Het onderzoek bestudeerde de trainingsintensiteit verdeling voor renners gedurende 7 maanden. Daarin was een duidelijke trend te merken van een Base periode (opbouw) naar een Pyramidal intensiteitsverdeling juist voor en tijdens competitie.

Of de Polarized- of Pyramidal-benadering nu de meest optimale intensiteitsverdeling is, blijft onderwerp van debat binnen de literatuur. Verschillende onderzoeken [8]–[10] suggereren dat de Polarized aanpak als de beste naar voren komt. Belangrijk is het om hierbij op te merken dat deze onderzoeken geen directe vergelijking maakten met de Pyramidal aanpak. Andere studies [10], [11], die de intensiteitsverdeling bij professionele atleten bestuderen, komen tot de conclusie dat deze gebruik maken van een Pyramidal verdeling. Daarnaast is er een studie van een Ironman 70.3 [7] die de aankomsttijden van atleten met een Pyramidal en een Polarized training aanpak met elkaar vergelijkte. In totaal werden 18 recreanten verdeeld over de 2 groepen. Na de wedstrijd bleek dat er slechts twee seconden verschil was tussen beide groepen voor geheel de Ironman 70.3. Dit verschil is veel te klein om te bewijzen dat de ene methode superieur zou zijn aan de andere methode.

Bovenstaande modellen geven een globaal overzicht van de manier waarop een atleet zijn training kan indelen, maar ze vertegenwoordigen een *one size fits all approach*. Daarnaast blijft een goede coach essentieel om de gekozen intensiteitsverdeling in een plan, aangepast aan de noden van de atleet, te verwerken. Een uitdaging voor de coach is dat niet elke atleet hetzelfde reageert op een trainingsprikkel [12]. Sommige atleten hebben juist meer baat bij een type van intervallen dan andere atleten. Om hierachter te komen zullen de coach en de atleet verschillende aanpakken samen moeten uitproberen en achteraf reflecteren of deze aanpak werkt voor de atleet op sportief en mentaal vlak. De studie van Bahensky [12] in 2021 deed meer onderzoek naar de personalisatie van een trainingsplan. Door middel van HRV-metingen (waarover meer uitleg volgt in een volgend hoofdstuk) voor een training werd bepaald hoe de vorige training verwerkt was. Op basis van deze gegevens werd besloten of de nieuwe training al dan niet aangepast moest worden. Elke deelnemer kreeg op deze manier toch een licht gepersonaliseerde aanpak voorgeschoteld. Zowel de controle als de HRV-groep maakte progressie in conditie. Maar opmerkelijk was dat iedereen van de HRV-groep verschillende persoonlijke records wist neer te zetten. Daarnaast was niemand van de HRV-groep ziek geworden. Zelfs geen enkele blessure of symptomen van overtraining waren op te merken binnen de groep. Dergelijk onderzoek doet vermoeden dat maatwerk in training op basis van HRV-metingen superieur is aan een *one size fits all* aanpak.

2.2 Stretchen

Naast sport specifieke trainingen is het voor atleten ook aangeraden om stretchen ook op te nemen in hun trainingsprogramma. Het vergroten van de flexibiliteit door stretching is positief gerelateerd aan een lagere kans op blessures [13]. Daarom wordt stretching voorgeschreven als een belangrijk onderdeel van de warming-up of cooling-down. Recentere studies duiden wel op het belang van de timing van stretchen. Hoewel stretchen gezond en aangeraden is, zou het niet langer meer voorgeschreven mogen in de warming-up van atleten die aanzienlijke kracht nodig hebben tijdens hun sport. Onderzoek heeft aangetoond dat de prestaties slechter waren wanneer stretching werd toegepast in vergelijking met situaties waarin proefpersonen helemaal geen warming-up deden. Wanneer rustig joggen als warming-up toegepast werd, verbeterden de prestaties aanzienlijk vooruit tegenover de baseline [14]. Andere onderzoeken met een direct verband met wielrennen bevestigen de bevindingen van het eerdere onderzoek. Bijvoorbeeld een onderzoek [15] in 2012 onderzocht de effecten van stretching op de *time to exhaustion* (TTE) en het zuurstofgebruik bij recreatieve fietsers die fietsen op 85% van hun maximale zuurstofopname (VO₂max). Dit onderzoek rapporteerde dat de recreanten die stretchen als opwarming hadden gebruikt tot 14 % meer zuurstof nodig hadden voor eenzelfde inspanning als de controlegroep. Daarnaast verminderde hun *time to exhaustion* met 26%.

Een ander onderzoek [16] toonde eveneens aan dat stretchen voor wielrennen een acute verlaging van de fietseconomie veroorzaakt. Fietseconomie kan worden beschreven als de relatie tussen de mechanische energie die door de fiets wordt gegenereerd en de chemische energie, afkomstig van het lichaam van de atleet, die nodig is om deze mechanische energie op te wekken. Met een acute verlaging van de fietseconomie wordt bedoeld dat een atleet tijdelijk meer energie moet leveren om een bepaalde output te verkrijgen. De auteurs raden daarom aan om statisch stretchen niet te gebruiken in een warming-up voor een fietstocht van gemiddelde intensiteit.

2.3 Krachttraining

Naast het verbeteren van de flexibiliteit, zou een atleet zijn absolute kracht kunnen verbeteren door krachttraining in zijn trainingsschema te verwerken. Volgens een systematisch review paper [17] zijn de resultaten voor krachttraining voor wielrennen uit de literatuur tot 2010 wisselend. Auteurs kwamen tot de conclusie dat willekeurige krachttraining bovenop het trainingprogramma van een atleet extra vermoeidheid veroorzaakt zonder een specifieke trainingsprikkel te geven. Hierbij zou het beoogde doel mogelijks niet bereikt worden. Echter rapporteren de auteurs dat wanneer een deel van de normale trainingen vervangen worden door krachttrainingen, er positieve resultaten kunnen ontstaan.

Er zijn verschillende onderzoeken terug te vinden vanaf 2010 waarbij de effectiviteit van krachttraining bewezen wordt. Zo rapporteert een onderzoek van 2017 significante verbeteringen: 11.2% toename in explosiviteit en 8.5% toename bij het VO₂max [18]. Een ander onderzoek [19] bevestigt de positieve bevindingen van bovenstaand onderzoek. Het onderzoek toonde aan dat de fietseconomie, de relatie tussen de mechanische output en de chemische energie (van het lichaam) nodig om dit te bereiken, bij 70% van het VO₂max een verbetering kende van 4.7%. Bovendien nam het vermogen dat de atleet kon trappen met 6.9% toe bij diezelfde inspanning.

Tevens is het relevant om een studie te beschouwen die zich richt op de inspanningen die veel amateurs op een zondagochtend leveren. Een voorbeeld hiervan is de studie door Rønnestad [20], waarin de impact van krachttraining werd onderzocht bij recreanten. In deze studie fietsten de deelnemers eerst 185 minuten op een rustig tempo, gevolgd door een maximale inspanning van vijf minuten aan het einde van de rit. Dergelijke inspanning is vergelijkbaar voor een recreatieve fietser die met zijn lokale fietsclub op zondagochtend twee tot drie uren fietst en tot slot met zijn allen een sprintje doet naar de lokale ontmoetingsplek op het einde van de rit. Het onderzoek toonde aan dat krachttraining het mediaan vermogen in de laatste vijf minuten kon verhogen met zeven procent. Deze bevinding suggereert dat na drie uur relatief rustig fietsen, positieve effecten van krachttraining nog steeds merkbaar zijn tijdens zware inspanningen.

Naast het vergroten van de beenkracht, wilt een recreatieve wielrenner deze nieuw verworven kracht ook behouden tijdens het seizoen. Een onderzoek door Rønnestad [21] toonde aan dat de verworven beenkracht over een periode van twaalf weken, waarin tweemaal per week krachttraining werd uitgevoerd, behouden kon blijven gedurende een periode van dertien weken met slechts één krachtsessie per week. Deze bevindingen suggereren dat regelmatige krachttraining gedurende het seizoen kan bijdragen aan het behoud van reeds verworven beenkrachten.

Naast trainingsfrequentie is de ideale rustperiode tussen verschillende soorten workouts van belang voor sporters die optimale verbeteringen willen. Zo is er een studie [22] die de combinatie tussen aerobe en krachttraining onderzocht met als conclusie dat atleten voldoende rust moeten voorzien tussen twee soorten workouts als ze optimale resultaten nastreven. Concreet betekent dit voor sporters dat het afgeraden wordt om op eenzelfde dag zowel aerobe als krachttraining te voorzien. Volgens het onderzoek zou er 24 uur of meer rust tussen de verschillende soorten trainingen moeten zijn om maximale neuromusculaire en oxidatieve aanpassingen te verkrijgen [22].

2.4 Energiesystemen voor een duursporter

De eenvoudigste vorm van energie die het lichaam nodig heeft voor alle processen is adenosinetrifosfaat, ook wel bekend als ATP. Dit molecuul is essentieel voor vitale processen en voor de voortbeweging van het lichaam. Het menselijk lichaam heeft echter slechts 80 tot 100 gram ATP opgeslagen als reserves [23]. Ter illustratie: het lopen van een complete marathon kost ongeveer één kilogram ATP [24]. Het menselijk lichaam moet dus voortdurend ATP produceren. Hiervoor kan het lichaam terugvallen op drie verschillende energiesystemen: het ATP-CP-systeem, het glycolitische systeem en het aerobe systeem. De eerste twee genoemde systemen kunnen op korte termijn een grote hoeveelheid ATP produceren, maar bij maximale inspanning kan slechts gedurende één minuut energie worden geleverd [23]. Wanneer het lichaam enkele minuten heeft gerust, kan dit systeem opnieuw worden aangesproken voor maximale inspanning. De eerder genoemde systemen zijn anaërobe systemen, wat betekent dat ze geen zuurstof vereisen voor de energieproductie. Voor duursporters zijn deze systemen minder relevant. Ze zouden bijvoorbeeld wel snel een korte heuvel kunnen beklimmen, maar dit zou geen verbetering opleveren tijdens een lange toertocht op de fiets. Als we de bijdrage van het aerobe en het anaerobe systeem tijdens een uur constant fietsen bekijken, wordt opgemerkt dat het aerobe systeem 98% van alle energie levert [25]. Binnen het aerobe systeem kan een verschil worden opgemerkt in de brandstof die wordt gebruikt voor het aanmaken van ATP. Het lichaam kan ervoor kiezen om koolhydraten of lichaamsvet te gebruiken. Koolhydraten in het aerobe systeem vereisen minder zuurstof om te worden omgezet in ATP, waardoor ze geschikt zijn voor intensievere inspanningen. Een belangrijk nadeel is echter dat de voorraad koolhydraten beperkt is in het lichaam, waardoor de recreatieve sporter deze voorraad tijdens het sporten moet aanvullen om de intensiteit te kunnen behouden. In sectie 2.4.1 is meer informatie te vinden over het juiste gebruik van koolhydraten tijdens lange duurinspanningen. Het aerobe systeem dat lichaamsvet gebruikt, vult het gebruik van koolhydraten perfect aan. Hoewel er meer zuurstof nodig is om lichaamsvet om te zetten naar ATP, zijn de reserves zuurstof en lichaamsvet onbeperkt voor duurinspanningen. Belangrijk om op te merken is dat het aerobe systeem beide brandstofbronnen tegelijkertijd gebruikt; de intensiteit van de inspanning bepaalt de verhouding tussen beide brandstoffen [26].

2.4.1 Koolhydraatoptimalisaties

Voor meer ervaren sporters die niet alleen hun uitdaging willen voltooien maar ook hun prestatietijd willen verbeteren, zijn verschillende technieken beschikbaar zowel voor als tijdens de inspanning rond het gebruik van koolhydraten. Een eerste stap naar het verbeteren van prestaties, is het toepassen van stapelen van koolhydraten, ook *carboloading* genaamd. Het doel van deze techniek is, zoals de naam reeds doet vermoeden, extra koolhydraten in de spieren als glycogeen opslaan om zo een prestatiewinst te boeken. Deze techniek dient een aantal dagen voor de belangrijke dag gebruikt te worden. Ten eerste moet de atleet ervoor zorgen dat zoveel mogelijk van de glycogeenreserves van het lichaam opgebruikt zijn. Vervolgens mag de atleet in de dagen voor het evenement meer en meer koolhydraten eten om zijn reserves aan te vullen. Op de laatste dag van de voorbereiding mag hij zelfs tot tien gram koolhydraten per kilogram lichaamsgewicht consumeren om zo de glycogeenvoorraad te optimaliseren. Uit onderzoek [27] blijkt dat dit een prestatiewinst kan opleveren van twee tot drie procent wanneer het evenement langer duurt dan 90 minuten. Bij inspanningen korter dan 90 minuten, heeft deze techniek geen aantoonbaar effect. Belangrijk om op te merken is dat meer koolhydraten eten dan 10 gram per kilogram lichaamsgewicht nog voor een grotere glycogeenreserve kan zorgen, maar geen aantoonbare winst voor de prestatie. In sommige gevallen zou de prestatie zelfs kunnen verslechteren wegens een gewichtstoename. Eén gram koolhydraat houdt ongeveer drie gram vocht vast, wat in sommige gevallen niet wenselijk is voor de sporter.

Een tweede manier om prestaties te verbeteren, is het consumeren van koolhydraten tijdens de inspanning. Dit is niet verwonderlijk, gezien het lichaam behoefte heeft aan koolhydraten tijdens inspanningen. Uit onderzoek [27], uitgevoerd in een laboratoriumomgeving, blijkt dat 60

tot 80 gram koolhydraten per uur de ideale hoeveelheid zou zijn voor langdurige inspanningen. Een andere bron [28] raadt bijvoorbeeld aan om 90 gram koolhydraten per uur te gebruiken wanneer een complete Ironman moet worden voltooid. Uit eenzelfde onderzoek blijkt ook dat een snellere tijd gerelateerd is aan een grotere consumptie van koolhydraten tijdens de wedstrijd. Interessanter voor dit onderzoek is het om te kijken naar de hoeveelheid koolhydraten die atleten gebruiken tijdens het fietsgedeelte van een complete Ironman (180 km). Daaruit blijkt dat ze zelfs anderhalve gram koolhydraat per kilogram lichaamsgewicht gebruiken gedurende het fietsgedeelte. Deze cijfers zouden erop kunnen wijzen dat een inname rond de 100 gram koolhydraten per uur nog beter zou kunnen zijn in de praktijk dan eerdere resultaten uit een laboratoriumomgeving suggereren.

Voor een beginnende recreant zijn deze hoeveelheden echter niet aan te raden vanwege mogelijke maag- en darmklachten. Uit onderzoek blijkt dat het mogelijk is om de maag te trainen in het verwerken van koolhydraten [27]. Dit lijkt een aanrader voor atleten die maximale uithoudingsprestaties willen bereiken. Daarnaast is het belangrijk om een optimale combinatie van verschillende vormen van koolhydraten te gebruiken. Een ander onderzoek toont aan dat er niet meer dan 60 gram koolhydraat in de vorm van glucose opgenomen kan worden [29]. Om meer koolhydraten te verwerken, is het aanbevolen om naast glucose ook fructose te gebruiken. Het lichaam verwerkt glucose en fructose op verschillende manieren, waardoor een combinatie van beide vormen ervoor kan zorgen dat atleten meer koolhydraten kunnen opnemen.

Naast langdurige inspanning is aangetoond dat het gebruik van koolhydraten voor een zware inspanning van ongeveer één uur ook nog positieve effecten heeft. Dit effect zou niet zozeer voortkomen uit het metabolisme van een atleet, gezien het lichaam genoeg voorraad heeft voor een zware inspanning van een uur. Onderzoek doet vermoeden dat de positieve effecten vanuit het centrale zenuwstelsel zouden komen [27]. Tot deze conclusie zijn onderzoekers gekomen nadat was vastgesteld dat een mondspoeling met koolhydraten hetzelfde effect had als het drinken van sportdrank met koolhydraten. Binnen dit onderzoek wordt aangegeven dat verder onderzoek nodig is om correct te kunnen identificeren hoe de prestatieverbetering tot stand komt.

Naast voeding moeten atleten waakzaam zijn dat ze voldoende drinken om zo het vochtverlies door zweten te compenseren. Deze stap is zeer belangrijk voor een langdurig evenement in een warm klimaat. Een vochtverlies van ongeveer twee tot drie procent van het lichaamsgewicht is al voldoende om een reductie in prestatie te veroorzaken. Daarnaast zal ook de lichaamstemperatuur toenemen, samen met de waargenomen perceptie van de moeilijkheid van de inspanning [28].

2.5 Supplementen

Algemene voedingskwaliteit en een effectieve strategie voor voedingsinname, zowel tijdens als na een fietstocht, vormen de fundamentele basis voor optimale atletische prestaties. Professionele atleten bouwen echter verder op deze basis door het gebruik van supplementen om hun prestaties nog verder te verbeteren.

Cafeïne

Cafeïne is een stimulerende stof voor het centrale zenuwstelsel. Positieve effecten van cafeïne zijn het verbeteren van aandacht en het verminderen van vermoeidheid. Cafeïne wordt in het dagelijks leven genuttigd door het gebruik van koffie, thee of frisdrank. Volgens de Vlaamse Hoge Gezondheidsraad mag een gezonde volwassene maximaal 5.7mg/kg/dag consumeren. Voor een persoon van 60kg komt dit met 340mg cafeïne, wat ongeveer 4 kopjes koffie van 125ml zijn [30].

De eerder genoemde effecten van cafeïne zijn interessant voor sporters tijdens lange tochten of wedstrijden op de fiets. Naast deze effecten heeft onderzoek reeds aangetoond dat het gebruik

van cafeïne kan leiden tot een verhoogd vermogen bij sporters, en bijgevolg ook tot betere prestaties. Uit een overzicht van verschillende studies in het onderzoek van Burke [31] blijkt dat niet alle uitgevoerde experimenten de positieve effecten van cafeïne kunnen meten. Desalniettemin rapporteert het merendeel van de onderzoeken dat cafeïne de prestaties positief beïnvloedt, zowel bij langdurige inspanningen van meer dan 60 minuten als bij korte en intensieve inspanningen van minder dan 60 minuten.

Cafeïne kan zowel voor als tijdens inspanning worden geconsumeerd en positieve effecten bieden, op voorwaarde dat de dosis cafeïne groot genoeg is. Uit een studie [32] blijkt dat 1 mg/kg niet voldoende is om prestaties te verbeteren. Dezelfde studie rapporteert dat een dosis cafeïne groter dan of gelijk aan twee mg/kg zorgt voor prestatieverbeteringen met een grootte van drie tot vier procent in het vermogen van een sporter gedurende een inspanning van 15 minuten. De cafeïne werd voor de start van de activiteit ingenomen. Een dosis kleiner dan drie mg/kg wordt in de literatuur als een kleine dosis beschouwd. Een grote dosis heeft niet altijd een groter effect, maar er is wel een onderzoek met een zeer hoge dosis van 500 mg dat een verbetering van 7 procent in het vermogen van atleten tijdens twee uur fietsen liet zien. Het gebruik van cafeïne heeft niet alleen voordelen. Cafeïne zorgt ervoor dat vocht uit het lichaam wordt verdreven, waardoor het afgeraden wordt (veel) cafeïne te consumeren tijdens warme omstandigheden.

Creatine

Creatine [33] is een combinatie van aminozuren die een integraal onderdeel vormt van een gebalanceerd dieet. De stof wordt hoofdzakelijk aangetroffen in rood vlees, vis en schaaldieren. Ongeveer 95 procent van alle creatine in het menselijk lichaam bevindt zich in de spieren, terwijl de overige vijf procent in de hersenen aanwezig is. Ongeveer tweederde van de creatine in de spieren bestaat uit fosfaatcreatine, wat snel energie kan vrijmaken voor zowel spieren als hersenen. Het resterende deel van de creatine in het lichaam is vrije creatine.

Gemiddeld kan het lichaam ongeveer 120 mmol/kg vetvrije spiermassa aan creatine vasthouden, met een maximum van 160 mmol/kg vetvrije spiermassa. Daarnaast wordt dagelijks ongeveer 1 tot 2 procent van de creatine in de spieren omgezet tot creatinine, een bijproduct van de spierwerking dat via de urine het lichaam verlaat. Hierdoor moet de creatinevoorraad dagelijks worden aangevuld met één tot drie gram creatine per dag. Ter referentie, dit komt overeen met ongeveer een halve kilogram onbewerkt rundsvlees. Slechts de helft van de benodigde creatine moet rechtstreeks via voeding worden aangevuld; de andere helft kan door het lichaam zelf worden geproduceerd, voornamelijk in de lever en de nieren.

In een normaal dieet waarbij een persoon één tot twee gram creatine per dag aanvult, is de creatineopslag tussen de 60 en 80 procent verzadigd. Dit betekent dat er nog ruimte is voor 40 procent aanvullende creatinesupplementatie. Volgens dit onderzoek is een effectieve supplementatie-methode om gedurende 5 tot 7 dagen viermaal per dag vijf gram creatine in te nemen. De totale hoeveelheid gesupplementeerde creatine moet ruwweg 0.3g/kg lichaamsgewicht bedragen. Zodra de creatinevoorraad verzadigd is, is het voldoende om dagelijks drie tot vijf gram creatine in te nemen om de hoge concentratie creatine te onderhouden. Een dergelijke creatinesupplementatie biedt voordelen voor een sporter. Prestaties bij hoge intensiteit kunnen verbeteren met 10 tot 20 procent. In een ander onderzoek [34] met 23 mannelijke recreanten die vijf keer een sprint van 15 seconden moesten uitvoeren, met twee minuten rust ertussen, werd een verbetering van vijf procent voor gemiddeld vermogen en vier procent voor maximaal vermogen vastgesteld. Deze verbetering was mogelijk na een creatinesupplementatie gedurende 6 weken. Binnen het onderzoek werd gekozen voor een spurduur van 15 seconden omdat het lichaam dan vooral steunt op het fosfaat-creatine-energiesysteem om vermogen te ontwikkelen. Een sprint van lange duur, bijvoorbeeld ongeveer 30 seconden zou voor 75 procent afhankelijk zijn van glycogeen als brandstof.

Naast een directe verbetering in prestaties wordt creatinesupplementatie ook gelinkt aan een verbeterde spierontwikkeling en bijgevolg ook aan een verbeterde revalidatie na een letsel. Onderzoek [33] toonde aan dat creatine verantwoordelijk was voor een verbetering van tien procent

in de dwarsdoorsnede van de spiervezels en 25 procent maximale kracht gedurende de revalidatie, in vergelijking met een controlegroep die geen creatine kreeg. Een ander voordeel [33] van creatine is dat deze stof, in tegenstelling tot cafeïne, vocht in het lichaam vasthoudt. Creatinesupplementatie zou 0.5 tot 1 liter vocht vasthouden, wat deze strategie een voordeel geeft tijdens wedstrijden of evenementen in warme omgevingen.

Nitraat

Nitraat is een stof die in vrij hoge dosis te vinden is in het sap van rode bieten. Verschillende onderzoeken tonen aan dat het drinken van dit sap prestatiebevorderend kan werken. Niet elk onderzoek kan dit helemaal bevestigen; er zijn ook onderzoeken te vinden die geen verbetering in prestaties kunnen opmerken na het toedienen van bietensap aan hun deelnemers. In het onderzoek [35] moest een groep deelnemers een tijdrif van 30 minuten rijden. Gedurende zeven dagen moesten de deelnemers een supplement nemen om nitraatgehaltes te verhogen. Na de supplementatie werden de tests afgenomen, gevolgd door een week rust zonder supplementatie. Daarna werd weer een week gesupplementeerd en werden opnieuw tests afgenomen. Deelnemers wisten niet in welke week ze effectief een nitraathoudende drank kregen en wanneer ze een placebo kregen. Gemiddeld genomen verbeterde het vermogen van de deelnemers met 2,8 procent bij het gebruik van nitraat ten opzichte van het placebo.

Bovenstaand onderzoek maakt gebruik van een langdurige methode voor de nitraatsupplementatie. Er zijn echter ook onderzoeken die aantonen dat slechts een dosis voor het sporten voldoende kan zijn om aantoonbare effecten waar te nemen. Zo is er het onderzoek van Muggerridge [36], waarbij deelnemers drie uur voor de inspanning een drankje moesten drinken. Ook binnen dit onderzoek werd er gewerkt met bietensap en bietensap zonder nitraat om het placebo-effect tegen te gaan. Deelnemers moesten fietsen op een gesimuleerde hoogte van 2500m en een tijdrif van 16 km afleggen. Door het drinken van bietensap met nitraat verbeterden de deelnemers hun tijd met 2,9 procent. Een soortgelijk onderzoek [37] met een tijdrif van 4 km en 16 km, niet op hoogte gesimuleerd, rapporteerde respectievelijk een verbetering van de tijd van 2,8 procent en 2,7 procent. Een ander soortgelijk onderzoek [38], waarbij verbeteringen werden gezocht rond lactaatgrenzen LT1 en LT2, kon in beide tests geen significante verbeteringen vinden. Daarnaast werd er ook onderzoek gedaan naar de effecten van nitraatrijk bietensap op langdurige inspanning. In het onderzoek van Wilkerson [39] moesten deelnemers een tijdrif afleggen van ongeveer 80 kilometer. Hierbij werd weer gebruik gemaakt van acute supplementatie met bietensap twee tot drie uur voor de inspanning. In tegenstelling tot bovenstaande onderzoeken werden voor dit onderzoek zeer goed getrainde fietsers gerekruteerd. Ze deden gemiddeld ongeveer zes trainingen per week, goed voor gemiddeld 11 uur training per week. Uit dit onderzoek blijkt dat nitraatrijk bietensap een verbetering van de tijd met 0,8 procent opleverde. De auteurs merken echter op dat het onderzoek zelf geen significant resultaat oplevert. Interessanter is hoe het resultaat tot stand komt. Binnen dit onderzoek wordt opgemerkt dat er mensen zijn die niet reageren op nitraatrijk bietensap. Als deze kleine groep buiten beschouwing wordt gelaten, wordt een verbetering van de tijd van twee procent vastgesteld. Hoewel er geen significante resultaten zijn, kan dit erop wijzen dat ook tijdens langdurige inspanningen het gebruik van nitraatrijk bietensap voor sommige atleten een voordeel kan opleveren.

Bicarbonaat

Bij een hoge intensiteit van 80 tot 100 procent van de maximale hartslag [40] wordt een accumulatie van waterstofkationen (positief geladen waterstofdeeltjes) geassocieerd binnen in het cytosol. Het cytosol is de vloeistof waarin metabole processen kunnen plaatsvinden en waarin de celorganen zich bevinden. Bij een overvloed aan waterstofkationen (ook wel acidose genoemd) daalt de intramusculaire pH, wat als gevolg heeft dat de glycolytische energieproductie vermindert. Bijgevolg kan een sporter minder energie leveren, wat als resultaat heeft dat de prestaties verminderen. Een verwerkingsmechanisme van het lichaam is om een voorraad koolzuur (H_2CO_3) te ontbinden naar HCO_3^- waaraan de H^+ ionen zich kunnen binden.

Van dit verwerkingsysteem maakt de supplementatie van bicarbonaat gebruik. Bicarbonaat zorgt voor een toename van HCO_3^- in het lichaam, waardoor er tijdens inspanning meer waterstofkationen kunnen worden aangemaakt, waardoor de inspanning langer kan duren.

2.6 Herstel

Net zo belangrijk als een stevige trainingsprikkel, is de kans voor het lichaam om te herstellen en beter te worden door supercompensatie. Dit fenomeen komt voor wanneer een lichaam zwaarder belast wordt dan dat het gewoon is. Deze belasting zorgt ervoor dat er een trainingsaanpassing mogelijk gemaakt kan worden. Het mechanisme hierachter wordt supercompensatie genoemd. De vermoeide spier herstelt na de inspanning tot een spier met een grotere kracht. Dit maakt het mogelijk voor een atleet om daarna te trainen met een intensiteit die daarvoor niet mogelijk was [41].

De aanpassing in de spier blijft niet permanent. Om gebruik te maken van de supercompensatie is het belangrijk dat de atleet op het juiste moment weer traint. Wacht de atleet te lang, dan heeft de vorige training geen nut meer. Traint de atleet te vroeg, dan zal hij last hebben van verminderde prestaties wegens onvoldoende herstel. Op termijn loopt hij zelfs het risico op overtraining, omdat hij meer van het lichaam vraagt dan het aankan. Deze effecten worden geschetst door figuur 2.1 [42].



Figuur 2.1: Effect van trainen bij onvoldoende herstel



Figuur 2.2: Effect van trainen bij optimaal herstel



Figuur 2.3: Effect training na teveel herstel

Het is aanbevolen dat een atleet traint wanneer hij volledig hersteld is. In figuur 2.2 is het effect van een correcte combinatie van training en herstel te zien. Elke training draagt bij tot het verbeteren van de prestaties op lange termijn [42]. Echter, indien een atleet te lang wacht met trainen, kan de conditie stagneren of zelfs achteruit gaan. Dit fenomeen is weergegeven in figuur 2.3.

Naast optimaal trainen draagt een evenwichtig voedingspatroon bij tot het verder verbeteren van de conditie. Als (recreatieve) atleet is het aan te raden om het eiwitgehalte in je dieet in de gaten te houden. Eiwitten zijn namelijk de bouwstenen van het spierherstel na een zware training. Recreatieve sporters wordt daarom aangeraden om vijftien tot twintig procent van hun dagelijkse energiebehoefte uit eiwitten te halen [43], [44]. Uit onderzoek blijkt verder dat het moment waarop eiwitten geconsumeerd worden ook een belangrijke rol speelt. Om de kwaliteit van krachttraining of een andere hoog intensieve training te maximaliseren, wordt aangeraden om 35 gram koolhydraten samen met tien tot twintig gram eiwit te consumeren een uur voordat de training begint. Een optimale recovery shake zou ongeveer 1,1 gram koolhydraten samen met een gelijke hoeveelheid eiwit moeten bevatten en binnen 30 minuten na het voltooien van de training geconsumeerd moeten worden. Het is zeer belangrijk voor een atleet om een positieve eiwitbalans te behouden. Op deze manier krijgt een atleet meer eiwit binnen dan dat het lichaam nodig heeft voor herstel, wat een basisvereiste is om spieren te ontwikkelen [44].

Naast het gebruik van eiwit, bestaan in het professionele circuit tal van manieren om het herstelproces van het lichaam te versnellen, waaronder massage, actieve recovery, cryotherapie, koudwaterimmersie, hyperbare zuurstoftherapie, compressiekleding, elektrostimulatie, en meer. Dit onderzoek legt de nadruk op het verbeteren van prestaties bij amateurs, waarbij de financiële haalbaarheid van sommige technieken mogelijk onrealistisch lijkt voor amateurs. Bijgevolg zullen alleen actief herstel, passief herstel en herstel met massage met elkaar worden vergeleken, om recreanten een beter beeld te geven van een optimale herstelstrategie die ook haalbaar is voor hen.

Uit onderzoek [45] blijkt dat lichte beweging het meest effectief is voor herstel tussen sets van oefeningen. In eerste instantie werd een baseline bepaald door de proefpersonen. Vervolgens werden de proefpersonen willekeurig in groepen ingedeeld. De groep met licht actief herstel tussen de sets van oefeningen kon telkens hun baseline evenaren bij elke oefening. Dit was niet mogelijk met de passieve herstelmethode waar de proefpersonen significant minder kracht konden leveren naarmate ze meer oefeningen gedaan hadden. Twee andere onderzoeken [46], [47] tonen verdere verdeeldheid in de literatuur met betrekking tot massages als geschikte techniek voor het bevorderen van het herstel van een atleet. Onderzoek [47] suggereert dat de trainers van topsporters mogelijk eerst beter geïnformeerd moeten worden over wanneer en hoe ze massage moeten toepassen bij hun atleten. Verder wordt opgemerkt dat een groot deel van het bewijsmateriaal ten gunste van massages afkomstig is van anekdotes van professionele atleten zelf en niet altijd van gegevens die in een ideale laboratoriumomgeving zijn verzameld.

Bovenstaande paragraaf behandelde de meest belangrijke technieken voor dagelijkse herstelmaatregelen. Echter is het herstel tijdens de nacht nog belangrijker. Een goede nachtrust is van grote waarde voor optimaal herstel van elke atleet. Zelfs voor alle mensen wordt voldoende slaap sterk aanbevolen. Minstens 7 uur per nacht wordt geadviseerd voor een goede gezondheid [48]. Uit ander onderzoek [49] blijkt echter dat atleten extra behoefte hebben aan een goede nachtrust. Zo werd ontdekt dat een kortere nacht de prestaties van atleten met drie procent kon verminderen tijdens een tijdrit van 60 minuten. Wanneer atleten echter meer sliepen dan de *baseline*, verbeterden hun prestaties zelfs met drie procent gedurende een tijdrit van 60 minuten. Slaap lijkt dus een belangrijke rol te spelen bij zowel herstel als dagelijkse prestaties. Deze bevindingen worden bevestigd door een ander onderzoek [50]. Dit onderzoek onderzocht de effecten van langer slapen op basketbalspelers gedurende 5 tot 7 weken. Alle deelnemers moesten minstens tien uur in bed per dag doorbrengen. De nauwkeurigheid van de worpen verbeterde met negen procent. Daarnaast werden er ook negen procent meer driepunters gemeten. Naast deze objectieve metingen rapporteerden de deelnemers ook dat hun humeur opmerkelijk beter was nu ze meer tijd in bed doorbrachten.

Naast directe invloed op de prestaties van een atleet heeft slaap ook invloed op blessures bij atleten. Onderzoek [51] heeft reeds aangetoond dat een atleet die minder dan acht uur slaap per nacht krijgt, tot 1,7 keer meer kans heeft op een significante blessure dan een atleet die langer dan acht uur per nacht slaapt. De auteurs melden hierbij dat ze de precieze oorzaak niet kunnen achterhalen. Ze wijzen wel op de negatieve effecten van een kortere nacht op de reactietijd en cognitieve functies, waardoor ze deze als de meest waarschijnlijke factoren erkennen voor het veroorzaken van blessures.

Slaap heeft een indirecte invloed op het lichaamsgewicht van een atleet, net zoals blessures. Kortere nachten worden geassocieerd met veranderingen aan de hormonen in het metabolische systeem die gerelateerd zijn aan eetlust en voedselconsumptie [52]. Deze veranderingen kunnen een mogelijke verklaring zijn waarom mensen na een korte nacht sneller geneigd zijn om meer koolhydraten en tussendoortjes te consumeren [53]. Ook nemen mensen sneller een grotere portie dan voorheen [54]. Daarnaast blijkt dat de glucoseverwerking van het lichaam afneemt wanneer de atleet onvoldoende heeft geslapen [55]. Uit bovenstaande feiten kan geconcludeerd worden dat slaap belangrijk is voor de metabole gezondheid en het lichaamsgewicht van een atleet. Naast de sterke invloed op het gewicht zou slaap een beschermende rol kunnen spelen tegen ziekten. Uit onderzoek [56] blijkt dat mensen die minder dan 7 uur slapen tot 3 keer meer kans hebben om een verkoudheid te ontwikkelen als ze eraan worden blootgesteld, in vergelijking met mensen die meer dan acht uur slapen. Een ander onderzoek [57], dat voortbouwt op dit onderzoek, rapporteert verder dat het risico om een verkoudheid te ontwikkelen verder toeneemt naarmate de nachtrust korter wordt. In hun onderzoek werd aangetoond dat de kans op een verkoudheid significant toenam wanneer de nachtrust korter was dan zes uur, in vergelijking met een nachtrust van zeven uur of meer.

Tot slot heeft onderzoek reeds aangetoond dat slaap een sterke relatie heeft met het reguleren van pijnprikkels in het menselijk lichaam. Zo is er een onderzoek [58] dat een daling van acht procent van de pijntolerantiegrens heeft vastgesteld na slechts één slechte nachtrust. In contrast hiermee heeft een ander onderzoek [59] aangetoond dat het verlengen van de slaap gedurende vier nachten de pijngrens kan verhogen met maximaal 20 procent.

2.7 Bestaande applicaties

Applicaties en platformen om de gezondheid van een atleet te monitoren bestaan reeds en worden wereldwijd gebruikt. Het aantal smartdevices dat te koop is op de markt is bijna eindeloos, daarom worden enkel Garmin⁴ en Whoop⁵ als directe concurrenten beschreven in dit werk. Garmin werd gekozen vanwege hun specialisatie in sport en elektronica. Dit merk produceert fiets-GPS'en, hartslagmeters, indoor trainers en smartwatches. Smartwatches zijn voor dit werk het meest interessant vanwege de mogelijkheid om de gezondheid gedurende de hele dag te monitoren. Naast Garmin is er ook gekozen voor Whoop. Dit merk heeft een nicheproduct ontwikkeld dat zich richt op gebruikers die zeer bedreven zijn in het analyseren van sportdata. De applicatie biedt gedetailleerde statistieken en inzichten, maar dit komt met een prijs. De aanschaf van een apparaat (armband) kost ongeveer 50 euro, terwijl het abonnement, noodzakelijk voor toegang tot de app, momenteel 264 euro per jaar bedraagt. Vanwege de hoge en terugkerende kosten is dit product niet voor elke amateur geschikt.

2.7.1 Garmin horloges

Het merk Garmin heeft een uitgebreid assortiment smartwatches ontwikkeld, specifiek gericht op sporters. Voor diverse soorten sporters en verschillende budgetten biedt Garmin een breed scala aan wearables⁶. De verwachte functies van elke smartwatch zijn natuurlijk ook aanwezig op de

⁴<https://www.garmin.com/nl-BE/>

⁵<https://www.whoop.com/eu/en/>

⁶<https://www.garmin.com/en-US/c/wearables-smartwatches/>

smartwatches van Garmin. Denk bijvoorbeeld aan het bijhouden van je slaap, het monitoren van stressniveaus in het lichaam en het volgen van je hartslag gedurende de hele dag⁷.

Het voordeel van het gebruik van het Garmin-platform is dat dit platform deze bekende functies gebruikt om nieuwe concepten te introduceren, zoals *Body Battery*, die ook zeer relevant kunnen zijn voor deze masterproef. Met *Body Battery*⁸ kan je smartwatch aangeven hoeveel energie je hebt verbruikt en hoeveel je op dat moment nog over hebt. Het is te verwachten dat naarmate de dag vordert, de batterij steeds leger raakt. Garmin geeft ook aan dat boeiende ervaringen, activiteiten en workouts essentieel zijn voor een gezonde levensstijl. Dus als een gebruiker nog voldoende energie heeft, moet hij zeker niet aarzelen om deze te gebruiken. Bij een laag energieniveau is het eerder aan te raden om rustiger aan te doen om jezelf niet te overbelasten op die dag. De enige manier om de batterij op te laden, is door een goede nachtrust te hebben. Eventueel kan een goed getimed dutje ook voor een kleine oplaadbeurt zorgen.”

Body Battery is een momentopname van het lichaam van de atleet op een specifieke dag. Om meer rekening te houden met alle trainingen die een atleet heeft gedaan en nog moet doen, is *Training Readiness*⁹ ontwikkeld. Deze ontwikkeling geeft precies aan hoe goed een atleet hersteld is van zijn training. Een lage score betekent niet meteen dat een atleet niet zou mogen trainen. Het wordt aanbevolen om verder te kijken in zijn eigen schema om te bepalen of hij al dan niet moet trainen. De score voor *Training Readiness* is vooral gebaseerd op de nachtrust van zowel de afgelopen nacht als de voorgaande dagen, om eventueel slaapttekort in overweging te kunnen nemen. Daarnaast wordt de HRV gemeten om te controleren of het lichaam klaar is voor de training. Ten slotte wordt er ook rekening gehouden met de stressniveaus van de afgelopen dagen, omdat stress op een bepaalde manier verantwoordelijk is voor het verbruiken van energie. Naast fysieke kracht kan het mentale aspect ook erg belangrijk zijn tijdens het trainen.

2.7.2 Whoop

In tegenstelling tot Garmin maakt Whoop geen smartwatches, maar smart fitness trackers. Het is ook een kleiner bedrijf dat zich slechts op één enkel model richt om de gezondheid op te volgen. De kracht van Whoop ligt in de uitgebreide data-analyse, die naast de verzameling van data, plaatsvindt. Net als bij Garmin-apparaten wordt ook de slaap bijgehouden, zodat een atleet precies weet hoeveel minuten hij in bed heeft gelegen, heeft geslapen en welk percentage van zijn slaap hij in elke slaapfase heeft doorgebracht. Een unieke functie van Whoop is dat het in staat is om een schatting te maken van hoe lang de atleet moet slapen om voldoende hersteld te zijn na zijn training. Het algoritme houdt ook rekening met eventueel slaapttekort of dutjes die de atleet gedurende de dag heeft gedaan om slaapttekort weg te werken. Om de training van een atleet beter in kaart te brengen gedurende een dag, introduceert Whoop de term *Strain*¹⁰. Dit is een score tussen 0 en 21 die rekening houdt met verschillende parameters zoals de duur en intensiteit van de training, maar ook met stress die het lichaam gedurende de dag ervaart, bijvoorbeeld door stressvolle situaties in de werk- of thuisomgeving. Tot slot werkt Whoop ook met een *Recovery* score¹¹ om aan te geven hoe goed een atleet hersteld is en hoe klaar hij is voor zijn volgende training. Deze score is vergelijkbaar met de *Training Readiness* score van Garmin. Whoop geeft aan dat HRV de grootste invloed heeft op een positief resultaat voor herstel. Daarnaast wordt ook rekening gehouden met de volgende aspecten: slaap, rusthartslag, zuurstofsaturatie in het bloed en huidtemperatuur.

⁷<https://www.garmin.com/en-US/garmin-technology/health-science/>

⁸<https://www.garmin.com/en-US/garmin-technology/health-science/body-battery/>

⁹<https://www.garmin.com/en-US/garmin-technology/running-science/physiological-measurements/training-readiness/>

¹⁰<https://www.whoop.com/us/en/thelocker/how-does-whoop-strain-work-101/>

¹¹<https://www.whoop.com/us/en/thelocker/how-does-whoop-recovery-work-101/>

2.8 Bepaling van fysiologische stress van een sportactiviteit

Het bepalen van de fysiologische stress op het lichaam na een training is cruciaal. Het doel van atleten en hun coaches is namelijk om de best mogelijke prestaties te leveren voor een specifiek doel. Daarom is het essentieel dat de topvorm van de atleet samenvalt met de datum van het beoogde doel. Om dit te bereiken, is het noodzakelijk om trainingen en de bijbehorende fysiologische stress te monitoren, zodat de respons van het lichaam achteraf ook gemeten kan worden. Afhankelijk van de respons van een atleet kan het trainingsplan verder geïndividualiseerd worden door zijn coach. Individualisatie kan verder verbeterd worden met aangepaste trainingsintensiteiten, zie sectie 2.1. Verder is het ook nuttig om de totale stress of trainingbelasting over een langere periode in het oog te houden. Zo kan een coach voorkomen dat een atleet teveel inspanningen levert met een mogelijke overtraining tot gevolg of geblesseerd zou geraken.

Om de trainingsbelasting te kunnen meten, wordt er in de meeste modellen gebruik gemaakt van de intensiteit van de training samen met de duratie. Het meten van de duratie is eenvoudig, dit is de totale tijd dat een sporter bewogen heeft in een trainingssessie. Voor het bepalen van de intensiteit van de training, kan er gebruik gemaakt worden van verschillende parameters. Bijvoorbeeld, het gevoel of de waargenomen inspanning kan worden geëvalueerd. Het classificeren van de inspanning kan aan de hand van de Borg-schaal worden uitgevoerd [60]. Zoals eerder aangehaald, heeft slaap een sterke invloed op de pijngrens van de atleet wat ervoor kan zorgen dat deze manier niet heel accuraat is om de trainingbelasting te bepalen. Om een meer objectieve trainingbelasting te bepalen, kan er gebruik gemaakt worden van een hartslagmeter. Hiervoor moet de atleet ofwel een externe borstband bezitten ofwel kan hij gebruik maken van de hartslagmetingen van een smartwatch. Vervolgens kan de TRIMP calculatie toegepast worden [61]. Deze methode gebruikt de gemiddelde hartslag van de activiteit en de totale duratie (in minuten). Om een trainingbelasting score te bekomen wordt de volgende formule gebruikt: $\text{duratie} \times \text{gemiddelde hartslag}$. Deze methode heeft als groot nadeel dat er geen onderscheid gemaakt kan worden tussen trainingen met intervallen of trainingen waarin een langere maar stabiele inspanning geleverd werd. Deze kunnen eenzelfde trainingbelasting opleveren, maar toch een groot verschil in gevoel voor de atleet geven. Een oplossing hiervoor is het gebruik van gepersonaliseerde hartslagzones om te meten hoeveel tijd een atleet in elke zone heeft doorgebracht. Hoe hoger de zone, hoe meer trainingsbelasting per tijdseenheid wordt gegenereerd. Trainingsplatforms zoals TrainingPeaks definiëren verschillende hartslagzones met bijbehorende trainingsbelastingen. Tijd doorgebracht in elke hartslagzone wordt vermenigvuldigd met de corresponderende trainingsbelasting, waarna de som wordt genomen om tot een uiteindelijke score te komen. Vanwege deze herdefinitie van trainingsbelasting zal de resulterende score niet meer direct vergelijkbaar zijn met de originele TRIMP-calculatie. Desondanks wordt op deze manier gewaarborgd dat intervallen tijdens de training beter worden weerspiegeld en dat de trainingsbelasting meer op maat is voor de atleet [62].

Hoewel hartslag objectiever kan zijn dan de waargenomen inspanning van een atleet, is deze methode toch niet helemaal waterdicht. Een hartslag voor een gegeven inspanning kan per dag lichtjes verschillen. Voor de meest objectieve inschatting van de trainingsbelasting is het gebruik van een vermogensmeter de beste optie. Bij het bepalen van trainingsbelasting met vermogen, wordt automatisch ook rekening gehouden met mogelijke intervallen in de training van de atleet.

Hoofdstuk 3

Machine learning

Dit hoofdstuk biedt een beknopte introductie tot machine learning. Het voorziet de lezers van een overzicht van de basisprincipes van machine learning en bespreekt enkele specifieke algoritmen die in de praktijk worden toegepast binnen de context van medische gegevens. Deze algoritme zijn verder ook relevant binnen deze masterproef om het herstel van de recreatieve atleet te voorspellen. Het aantal besproken algoritmen in deze masterproef is zeker niet uitputtend voor dit uitgebreide vakgebied. Tot slot worden ook enkele metrieken besproken om een machine learning model te evalueren.

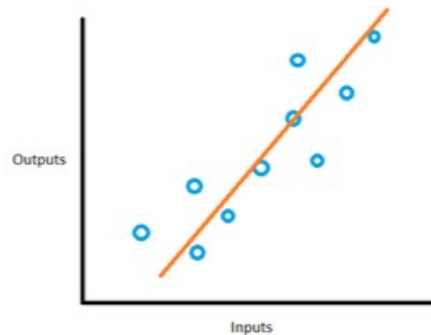
3.1 Wat is machine learning

Machine learning is een tak van de computerwetenschappen die tot doel heeft computers (algoritmen) kennis bij te brengen zonder dat deze specifiek geprogrammeerd worden. Binnen de computerwetenschappen valt machine learning onder de kunstmatige intelligentie, met een sterke focus op het behalen van duidelijke doelen of voorspellingen. Computers leren door ervaring op te doen. In de praktijk betekent dit het *fitten* van alle datapunten, wat je het beste visueel kunt voorstellen door alle datapunten in een assenstelsel te plotten. Vervolgens wordt een functie of curve berekend die een goede benadering vormt voor alle punten. Voor een visualisatie hiervan, zie figuur 3.1 De methode waarop deze functie of curve berekend wordt, hangt sterk af van het gekozen algoritme.

3.1.1 Onderverdelingen

Om een overzicht te bieden van de verschillende machine learning-algoritmen, hun werking en hun doelen, is het nuttig om ze te categoriseren en duidelijk te labelen.

Een fundamentele manier om machine learning-algoritmen te classificeren is op basis van het onderscheid tussen *supervised learning* en *unsupervised learning*. De term *supervised* verwijst naar de wijze waarop het algoritme tracht de relatie tussen de datapunten te modelleren. Bij *supervised learning* zijn de juiste uitkomsten (afhankelijke variabelen) voor elk datapunt in de dataset bekend. Na de trainingsfase kan de prestatie van het algoritme worden geëvalueerd aan de hand van deze bekende uitkomsten. Voorbeelden van prominente algoritmen binnen deze categorie zijn lineaire regressie, support vector machines en beslissingsbomen (*decision trees*). Bovendien kan de categorie *supervised learning* verder worden onderverdeeld in regressie en classificatie, afhankelijk van het type afhankelijke variabele. Wanneer de uitkomst in discrete categorieën moet worden ingedeeld, is classificatie de aangewezen methode. Daarentegen is regressie geschikt voor het voorspellen van een continue afhankelijke variabele. Deze onderverdeling helpt bij het kiezen van het juiste algoritme op basis van de aard van de voorspelde variabele [63].



Figuur 3.1: Fitten van curve tussen alle datapunten (bron: <https://www.quora.com/What-is-the-best-fitting-line-in-machine-learning>)

Anderszijds bestaat er ook *unsupervised learning*. Bij deze vorm van leren wordt geen juiste uitkomst van tevoren gedefinieerd; het is aan het algoritme om natuurlijke relaties en groepen binnen de dataset te identificeren. Bekende algoritmen binnen deze klasse zijn clusteringalgoritmen zoals k-means clustering en expectation-maximization clustering [63].

In deze masterproef zal een classificatiemodel worden gebruikt om uiteindelijk te beoordelen of een recreatieve atleet al dan niet voldoende is hersteld voor een volgende training. Daarom zullen de volgende secties zich specifiek richten op verschillende classificatiealgoritmen. Echter zijn sommige van deze algoritmen ook uitstekend geschikt zijn voor regressie.

3.1.2 Tree modellen

Binnen deze masterproef is er nood om een machine learning-algoritme te gebruiken zodat er een voorspelling gedaan kan worden voor het herstel voor morgen. In onderstaande paragrafen worden een aantal algoritmes opgelijst die als kandidaat kunnen gebruiken worden. Ook is deze lijst zeker niet eindig en zijn er nog andere valabele opties over om uit te kiezen.

De *Decision Tree* is een van de populairste en eenvoudigste algoritmen die er zijn. Het doel van dit algoritme is om een model te ontwikkelen voor de afhankelijke variabele op basis van een set beslissingsregels. Deze regels worden afgeleid door het algoritme zelf aan de hand van de kenmerken in de aangeleverde data. Bovendien heeft dit algoritme drie essentiële functies die kunnen worden gebruikt tijdens de constructie van het onderliggende model. Deze functies zijn: verdelen, stoppen en snoeien, waarbij de laatste in de literatuur ook bekend staat als *pruning*. Ten slotte zijn er nog twee componenten waarmee een boom kan worden opgebouwd: een tak (*branch*) en een knooppunt (*node*) [64].

Er bestaan drie soorten knooppunten. Als eerste is er het wortelknooppunt, ook bekend als de *root node*. Deze beslissing bepaalt in hoeveel exclusieve delen de dataset wordt verdeeld. Dit kan zijn in twee delen, maar eventueel kan deze ook in meer delen worden verdeeld. Omdat dit knooppunt als eerste komt, heeft deze nooit een input. Vervolgens zijn er de interne knooppunten. Deze geven aan welke mogelijkheden er zijn in de boomstructuur. De input van dit knooppunt komt altijd van een hoger gelegen knooppunt in de boom. Zijn eigen output zal altijd naar een lager gelegen knooppunt gaan. De lager gelegen knooppunten kunnen een blad zijn, maar dit is niet verplicht. De term 'blad' is in de literatuur ook gekend als een *leaf node*. Een blad geeft aan dat het einde van de boomstructuur bereikt is. De afgelegde weg naar het blad representeert de combinatie van beslissingen die zijn genomen om daar te komen [65].

De voordelen van het gebruik van een Decision Tree zijn de eenvoud waarmee beslissingen kunnen worden begrepen. Daarnaast is het een zeer robuust model dat niet erg gevoelig is voor uitschieters in de invoerwaarden. Een groot nadeel van het model is echter dat het sterk gevoelig is voor underfitting en overfitting bij kleine datasets [66]. Underfitting doet zich voor wanneer een model te eenvoudig is om alle complexe relaties tussen verschillende datapunten te vatten. Hierdoor vertoont het model aanzienlijke onnauwkeurigheid in zowel de trainings- als testdatasets. Aan de andere kant manifesteert overfitting zich wanneer een model uitstekende voorspellingen kan doen op de trainingsdataset, maar slecht presteert op de testdataset. Dit komt doordat het model te sterk is aangepast aan de trainingsdata en niet goed kan generaliseren naar niet eerder geziene datapunten.

Random Forest is een ander algoritme dat Decision Trees gebruikt en wordt toegepast op een breed scala aan classificatie- en regressieproblemen. Het algoritme creëert een groot aantal beslissingsbomen op basis van de gegeven dataset. Doordat het meerdere individuele modellen gebruikt, wordt Random Forest geclassificeerd als een ensemble-algoritme. Met meerdere beslissingsbomen moet het algoritme een methode bieden om een definitieve beslissing te nemen. Bij een bepaalde invoer genereert elke boom een uitvoer. Voor classificatie is de uiteindelijke uitvoer degene met de meerderheid van de stemmen van de bomen. Bij regressie wordt het gemiddelde genomen van de uitvoer van elke boom, aangezien regressieproblemen te maken hebben met continue waarden [67], [68].

3.1.3 Boosting modellen

Naast Random Forest maakt ook de boosting-familie gebruik van de ensemble-aanpak. Echter, bij een boosting-algoritme wordt geen gebruik gemaakt van het gemiddelde van alle trees, zoals bij Random Forest. Het idee achter boosting is om sequentieel nieuwe modellen te introduceren. Bij elke iteratie kan een nieuwe, zwakke basisleerder worden getraind op basis van de fouten van alle voorgaande algoritmes samen. Doordat er steeds meer focus wordt gelegd op een klein aantal fouten, kan het uiteindelijke algoritme erg nauwkeurige voorspellingen doen. Echter kan dit ook leiden tot sterke overfitting, zoals aangetoond in onderzoek waarin deze soort algoritmen alle andere geteste algoritmen konden verslaan of volledig niet toepasbaar waren vanwege de overfitting [69]. Een ander nadeel van een boosting-algoritme is dat deze over het algemeen trager is door de keuze voor hun ensemble-methode. Door sequentieel modellen te trainen, ligt de snelheid lager dan wanneer bijvoorbeeld een Random Forest wordt gebruikt, waarbij elk model parallel kan worden getraind [70], [71].

Een bekend voorbeeld van een boosting-algoritme is Adaptive Boosting, kortweg AdaBoost genoemd. AdaBoost is ontworpen om meerdere zwakke classifiers te combineren tot een sterke classifier. Dit wordt bereikt door de fouten van een zwakke classifier te corrigeren met behulp van een nieuwe zwakke classifier. Concreet wordt dit geïmplementeerd door gewichten te gebruiken bij elke iteratie van het trainingsproces. Correcte identificaties krijgen een lager gewicht, terwijl foute classificaties een hoger gewicht krijgen. Door deze hogere gewichten kan het algoritme onderscheiden waarop gefocust moet worden. In het geval van AdaBoost wordt er meer nadruk gelegd op de hogere gewichten [71].

Naast AdaBoost is Gradient Boosting ook een bekend algoritme dat vrij analoog aan AdaBoost te werk gaat. Een cruciaal verschil tussen beide algoritmen is de manier waarop de fouten worden gelabeld en behandeld. Gradient Boosting labelt geen rechtstreekse fouten. In plaats daarvan wordt een verschil berekend tussen de voorspelde waarde en de werkelijke waarde. Vervolgens probeert elke iteratie een nieuw model te fitten naar deze berekende verschillen om de fouten verder te minimaliseren. Uiteindelijk is de finale voorspelling een sommatie van de verschillen van alle zwakke classifiers [70].

Om Gradient Boosting verder te optimaliseren, werd XGBoost ontwikkeld. Dit staat voor Extreme Gradient Boosting, dat net zoals Gradient Boosting steunt op dezelfde eigenschappen van de boosting-algoritmes. Het is belangrijk om op te merken dat XGBoost ontwikkeld is om extreem krachtig, schaalbaar en overdraagbaar te zijn. In tegenstelling tot de standaard

gradient boosting, kan XGBoost deels parallel worden uitgevoerd. Hierbij worden de individuele bomen in het algoritme gemaakt door meerdere cores te gebruiken. Bovendien werd de data beter geordend, wat heeft geleid tot een vermindering van de tijd die nodig is om een bepaald stuk data te vinden. Deze aanpassingen zorgen voor een verkleining van de trainingstijd van het algoritme [72].

3.2 Machine learning in de praktijk

Machine learning kan in de praktijk voor talloze doeleinden worden gebruikt, waaronder toepassingen in de gezondheidszorg. Wat sterk verwant is met de gegevens die in deze masterproef verwerkt worden. Machine learning-modellen kunnen ondersteuning bieden bij het stellen van diagnoses voor aandoeningen zoals borstkanker[73], longkanker[74], en bij het analyseren van endoscopiebeelden[75] of beelden van de CT-scanner[76]. Kortom, machine learning haalt in de praktijk goede resultaten op gezondheidsdata waardoor het interessant is om concrete voorbeelden uit de praktijk te bestuderen. Problemen en oplossingen daarvan kunnen relevant zijn voor deze masterproef waarin ook met soortgelijke gezondheidsdata gewerkt wordt.

Het gebruik van machine learning in de gezondheidssector vereist grondige monitoring, vooral gezien de ongewenste implicaties van een onevenwichtige dataverdeling. Een dergelijke verdeling ontstaat wanneer er aanzienlijk meer datapunten van de ene klasse (de meerderheidsklasse) zijn dan van de andere klasse (de minderheidsklasse). In dergelijke scenario's zal het machine learning-algoritme aanzienlijk partijdig zijn ten gunste van de meerderheidsklasse, resulterend in een vertekend beeld van de nauwkeurigheid van het model. Dit kan leiden tot consistente misclassificaties van instanties van de minderheidsklasse als behorend tot de meerderheidsklasse[68]. Dit vormt ook een uitdaging binnen deze masterproef. Recreatieve atleten trainen doorgaans minder intensief dan professionele atleten. Daarom is de verwachting dat de dagen waarop een recreatieve atleet onvoldoende herstelt, relatief beperkt zullen zijn in vergelijking met alle gemeten dagen.

Ook in de praktijk wordt door onderzoekers gestreefd om zeer nauwkeurige voorspellingen te doen met datasets waarbij er een minderheidsklasse aanwezig is[77]. Het belang hiervan wordt onderstreept door het feit dat een precieze voorspelling van gezonde individuen een beperkte bijdrage levert aan de verbetering van de gezondheidszorg. Het is daarentegen van vitaal belang om mensen met ernstige aandoeningen, zoals borst- of longkanker, nauwkeurig te kunnen diagnosticeren. Bij de evaluatie van machine learning-algoritmen binnen deze context is het daarom cruciaal om de nauwkeurigheid niet louter te beschouwen als een betrouwbare indicator van prestaties op onevenwichtige datasets[68].

Voor deze masterproef is het interessant om te onderzoeken welke concrete methoden worden toegepast om nauwkeurigere voorspellingen te doen met onevenwichtige datasets in de praktijk. Zo lijken er dus twee stromingen van aanpak te bestaan. Zo kan er een dataniveau-aanpak en een algoritmeniveau-aanpak onderscheiden worden. Bij de dataniveau-aanpak wordt de verdeling van de klassen aangepast tijdens de voorverwerkingsfase, voorafgaand aan het trainen van een machine learning-algoritme op de dataset. In tegenstelling hiermee tracht de algoritmeniveau-aanpak het algoritme zelf aan te passen om beter om te kunnen gaan met de minderheidsklasse. Over het algemeen wordt in de literatuur vaak de voorkeur gegeven aan de dataniveau-aanpak. Deze voorkeur kan worden toegeschreven aan de grotere mate van controle die hiermee kan worden uitgeoefend over de balans in de dataset. Bovendien kan de balans achteraf nog eenvoudig worden aangepast door items toe te voegen aan of te verwijderen uit de dataset, afhankelijk van de specifieke situatie. Een andere overweging is dat de dataniveau-aanpak gemakkelijker te implementeren is en kan worden geïntegreerd met verschillende bestaande machine learning-algoritmen. In tegenstelling hiermee vereist de algoritmeniveau-aanpak aanpassingen aan bestaande algoritmen om rekening te houden met het onevenwicht in de klassen[78].

Uit de literatuur blijkt dat de dataniveau-aanpak gemakkelijker te implementeren is dan de algoritmeniveau-aanpak. Het is daarom interessanter om de dataniveau-aanpak verder te bestu-

deren voor eventuele toepassingen binnen deze masterproef. Ook binnen de dataniveau-aanpak kunnen twee benaderingen worden onderscheiden: over-sampling en under-sampling methoden. Bij under-sampling worden datapunten uit de meerderheidsklasse willekeurig verwijderd. Deze methode wordt echter niet aanbevolen voor kleine datasets, aangezien dit kan resulteren in een aanzienlijk verlies van data. In het geval van de over-sampling methode worden op synthetische wijze meer datapunten gegenereerd voor de minderheidsklasse. Het voordeel hiervan is dat er geen eerder verzamelde datapunten hoeven te worden verwijderd[79].

De over-sampling methode wordt vaker gekozen boven de andere aanpak. De over-sampling aanpak staat ook bekend als SMOTE (Synthetic Minority Over-sampling Technique)[80]. Met SMOTE wordt het onevenwicht in de dataset aangepakt door synthetische datapunten te genereren voor de minderheidsklasse, waarbij gebruik wordt gemaakt van de Euclidische afstand tussen verschillende reeds bestaande burenen. Het concept achter SMOTE is dat het nieuw gegenereerde datapunt gelijkaardig is aan de bestaande data, aangezien het wordt afgeleid van de kenmerken van de nabijgelegen geselecteerde punten. Een belangrijke kanttekening is echter dat het originele algoritme mogelijk niet de meest optimale keuze is voor datasets met veel dimensies, omdat het originele SMOTE-algoritme extra ruis aan de dataset kan toevoegen. Desondanks wordt SMOTE toch vaak toegepast in toepassingen binnen de medische sector waar klasse-onevenwicht een uitdaging vormt[81].

3.3 Verschillende evaluatie metrieken

Binnen het domein van machine learning zijn diverse evaluatiemetrieken beschikbaar om de prestaties van een model te beoordelen. De keuze voor een specifieke metriek hangt in eerste instantie af van de aard van het machine learning-probleem, zoals classificatie of regressie, waarvoor verschillende metrieken zijn ontwikkeld op basis van de eigenschappen van deze problemen. Daarnaast speelt het doel van de machine learning-toepassing een noodzakelijke rol: welk type voorspelling wordt gedaan en in welke context.

In sommige gevallen is het streven naar een hoge mate van nauwkeurigheid van cruciaal belang, zoals bij medische diagnoses of financiële voorspellingen. Echter, in andere scenario's, zoals bij het identificeren van phishing-e-mails, is niet alleen een hoge nauwkeurigheid van belang, maar ook het vermijden van foutieve classificaties van phishing-e-mails als legitiem, vanwege de mogelijke financiële en beveiligingsrisico's die daarmee gepaard gaan.

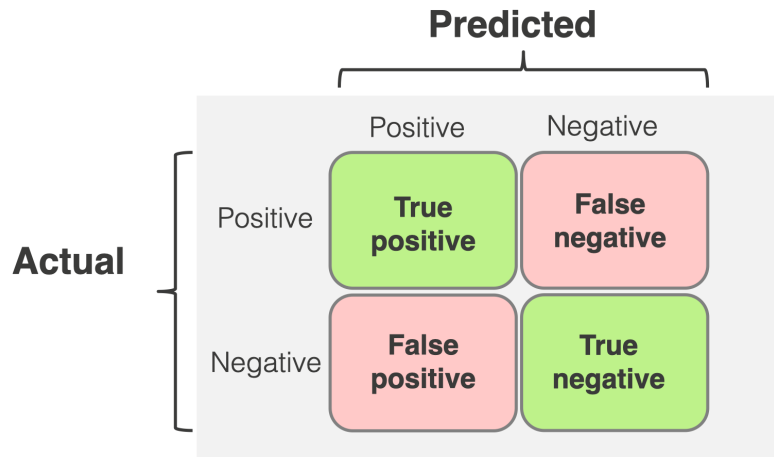
Bij het evalueren van een classificatiemodel zijn er vier fundamentele bouwstenen die altijd worden gebruikt: *true positives* (TP), *false positives* (FP), *false negatives* (FN) en *true negatives* (TN). *True positives* zijn de datapunten die correct als positief worden geïdentificeerd door het model. *False positives* zijn de datapunten die onterecht als positief worden geïdentificeerd. *False negatives* zijn de datapunten die ten onrechte als negatief worden geïdentificeerd door het model. Tot slot zijn er de *true negatives*, wat verwijst naar de datapunten die correct als negatief worden geïdentificeerd.

Door deze metrieken te combineren, kunnen we een confusion matrix construeren, zoals afgebeeld in figuur 3.2.

Bovenstaande metrieken bieden nuttige inzichten, maar geven niet direct een overzicht van de algehele prestatie van een model. Ook maken ze het moeilijk om verschillende modellen met verschillende datasets rechtstreeks met elkaar te vergelijken in één oogopslag. Daarom is het waardevol om op basis van deze bouwstenen nieuwe metrieken te berekenen die meer inzicht kunnen bieden in de prestaties van een model. Enkele van deze metrieken zijn:

Accuraatheid

De accuraatheid in classificatie is een van de meest fundamentele en toegankelijke evaluatiemetingen. Het wordt berekend door het aantal correcte voorspellingen (TP + TN) te delen door het totale aantal voorspellingen, en vervolgens te vermenigvuldigen met 100 om een per-



Figuur 3.2: Confusion Matrix 2x2 (bron: <https://www.evidentlyai.com/classification-metrics/confusion-matrix>)

centage te verkrijgen. Deze berekening kan worden uitgedrukt door middel van de volgende formule:

$$\text{Accuraatheid} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100 \quad (3.1)$$

Sensitiviteit

De sensitiviteit, ook wel *recall* genoemd in sommige literatuur, wordt gedefinieerd als de verhouding tussen alle *true positives* (TP) en alle positief gelabelde gevallen (TP + FP). Met andere woorden kan de sensitiviteit het beste worden omschreven als het aandeel correcte positieve voorspellingen ten opzichte van alle werkelijke positieve gevallen. Een hogere sensitiviteit betekent dat het model minder vals negatieve voorspellingen genereert, wat aangeeft dat het model effectief is in het identificeren van positieve gevallen. Een sensitiviteit die dicht bij 1 ligt, geeft aan dat het model weinig werkelijk *true positives* (TP) heeft gemist en dus in staat is om alle *true positives* correct te labelen. Een lage score (< 0.5) duidt erop dat het model veel *false negatives* (FN) labels genereert, wat afhankelijk van de context waarin de labels worden gebruikt, een probleem kan zijn dat moet worden aangepakt.

$$\text{Sensitiviteit (Recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.2)$$

Precisie

De precisie, ook bekend als *precision* in de Engelse literatuur, wordt gedefinieerd als de verhouding tussen de correct gelabelde waarden, de *true positives* (TP) en het totaal aantal positief gelabelde waarden (TP + FP). Met andere woorden kan precisie het beste omschreven worden als de mate van nauwkeurigheid van een positieve voorspelling van het model. Een hoge preciescore, dicht bij 1, geeft aan dat het model geen werkelijke *true positives* (TP) heeft gemist en goed onderscheid kan maken tussen correcte en onjuiste labeling. Het belangrijkste aspect van precisie is dat er weinig *false positives* zijn bij het labelen van de *true positives* (TP). Net als bij sensitiviteit kan een lage preciescore een probleem vormen in bepaalde contexten. Het is dan ook de context die zal bepalen welke van deze twee metrieken de voorkeur krijgt.

$$\text{Precisie} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3.3)$$

3.4 Kruisvalidatie

Kruisvalidatie is een veelvoorkomende methode in machine learning voor het evalueren van de prestaties van een model. Deze methode zorgt ervoor dat het model niet slechts één keer wordt getest met de training en testset, maar meerdere keren, wat een robuustere schatting van de prestaties van het model mogelijk maakt. Een veelgebruikte techniek voor kruisvalidatie is de *k-fold* methode. De parameter k in de naam duidt op het aantal segmenten waarin de dataset wordt verdeeld. Gebruikelijke waarden voor k zijn vijf of tien, afhankelijk van de grootte van de dataset en de behoefte aan voldoende trainings- en testdata in elke fold. Deze voorwaarde is meer van toepassing op kleinere datasets.

In het geval van deze masterproef, waar de dataset relatief klein is, is gekozen voor een *k-fold* methode met $k=5$. Deze keuze balanceert het aantal segmenten zodat de dataset voldoende wordt gesegmenteerd, terwijl elk segment groot genoeg blijft om betekenisvolle trainings- en testresultaten te leveren. Hieronder worden de stappen beschreven die de *k-fold* kruisvalidatie in de praktijk toepast:

1. Dataset shufflen zodat alle rijen willekeurig door elkaar zitten.
2. Splits in dataset in k delen (vijf voor deze masterproef).
3. Herhaal k keer.
 - (a) Kies een deel, die nog niet eerder als testset gebruikt werd, als testset.
 - (b) Gebruik de andere delen als trainingset.
 - (c) Kies een machine learning model en train vervolgens met de trainingset.
 - (d) Onthou de evaluatie score.
4. Neem het gemiddelde van alle evaluatie scores.

Hoofdstuk 4

Verzameling gegevens voor inschatting herstelstatus

In deze masterproef wordt gestreefd naar het optimaliseren van de prestaties van recreatieve wielrenners door middel van een verhoogde focus op het herstelproces van de atleten. Deze studie richt zich op het in kaart brengen van diverse aspecten van herstel in de dagelijkse context van recreatieve sporters. Het doel is om dit herstelproces te meten in realistische omstandigheden en de verkregen gegevens nadien te analyseren.

4.1 Datacollectie

Alvorens met de dataverzameling te starten, moest worden bepaald wat de doelgroep van deze studie zou zijn. De eerste vereiste was dat de deelnemers voldoende aan sport deden, zodat hun lichaam daadwerkelijk een herstelproces zou ondergaan. Er werden geen andere vereisten opgelegd bij de selectie van de proefpersonen. Uiteindelijk werden vijf mannelijke deelnemers geselecteerd, in de leeftijd van 18 tot 25 jaar. Deze deelnemers werden gevraagd om gedurende een periode van drie maanden een smartwatch te dragen. Tijdens de datacollectie werd gebruik gemaakt van een Fitbit-smartwatch van het type Sense 2. Zover bekend en op het moment van schrijven is dit het *flagship* van Fitbit¹.

4.1.1 Meetbare parameters

De volgende parameters zijn van belang voor het onderzoek. Deze kunnen zowel gemeten zijn als gebaseerd zijn op berekeningen op gemeten data:

- Stress Management Score
- Slaap score
- Tijd in verschillende slaapfases en totale slaap duratie
- Aantal stappen per dag
- Tijd gependeed sportend of zittend
- Hartslag Variabiliteit
- Hartslag in rust

Belangrijk om op te merken is dat de smartwatch nog meer parameters kan meten en extra functies heeft om inzicht te bieden in gezondheid of slaappatronen via de applicatie zelf. Echter,

¹<https://www.fitbit.com/global/be/products/smartwatches/sense2>

deze aspecten vallen buiten de scope van de masterproef. Hieronder worden de belangrijkste en complexere parameters uitgelegd. Het aantal stappen per dag, ademhaling, slaapduur en tijd doorgebracht in verschillende slaapfasen worden als vanzelfsprekend beschouwd en vereisen geen verdere uitleg.

Stress Management Score

De Stress Management Score is een meting die wordt berekend door Fitbit op basis van verschillende parameters. Deze score geeft een indicatie van hoe goed iemand op een bepaalde dag met stress kan omgaan. Een hogere score suggereert dat de persoon beter in staat is om met een stressvolle dag om te gaan. Een lagere score daarentegen duidt erop dat het raadzaam is voor de persoon om het rustiger aan te doen en zichzelf niet te veel te belasten. Fitbit hanteert een brede definitie van stress, waarbij ook fysieke stress als gevolg van inspanning wordt meegenomen. De parameters die worden gebruikt bij de berekening van de score omvatten onder andere gegevens over de hartslag gedurende de dag, slaapgegevens en informatie over lichamelijke activiteiten. Fitbit heeft niet alle factoren onthuld die een rol spelen bij de berekening van deze score, maar volgens informatie op hun blog worden tot tien verschillende factoren gemeten door de smartwatch en gebruikt om de score te bepalen. De uiteindelijke Stress Management Score varieert tussen 1 en 100.

Om meer inzicht te verschaffen in de score, worden alle factoren opgedeeld in drie verschillende categorieën: responsiveness, inspanningsbalans en slaap. De eerste categorie omvat gegevens over de hartslag en de metingen van de EDA-sensor. Een EDA-sensor kan relevant zijn om een inschatting te maken van de hoeveelheid stress die het lichaam ervaart op basis van de geleidbaarheid van de huid. De tweede categorie onderzoekt de impact van fysieke activiteiten op het stressniveau van de drager van het horloge. De concrete berekening hiervan wordt echter niet vrijgegeven. Tot slot omvat de laatste categorie de invloed van slaapduur en -kwaliteit op het stressniveau. Fitbit geeft wel inzicht in de verschillende gewichten per categorie. Zo krijgen responsiveness en slaap elk een gewicht van 30 bij de scoreberekening, terwijl inspanningsbalans lichtjes zwaarder weegt met een gewicht van 40. Samen vormen deze gewichten de maximale score van 100 voor dit component. Een concreet voorbeeld is wanneer een gebruiker een score behaalt van 24/30 voor slaap, 25/30 voor responsiviteit en 30/40 voor inspanningsbalans. Zijn totale Stress Management Score bedraagt dan $24 + 25 + 30 = 79$ van de maximale 100 punten.

Een laatste belangrijke opmerking is dat de Stress Management Score van Fitbit in deze masterproef zal worden hernoemd tot een readiness score. Deze terminologie is meer gangbaar onder sporters en geeft aan dat ze gereed zijn voor een nieuwe training. Bij de term Stress Management Score kan mogelijk te sterk de associatie worden gelegd met mentale stress op het werk, terwijl deze score juist het fysieke component omvat. Deze readiness score zal vervolgens ook bepalen of een recreatieve atleet voldoende hersteld is of niet.

Slaapscore

De slaapscore is opnieuw een score die door Fitbit wordt berekend aan de hand van verschillende andere factoren die gemeten kunnen worden door de smartwatch. Een goede kwaliteit slaap wordt geassocieerd met een optimaal fysiek herstel, wat van cruciaal belang is voor het verwerken van trainingen. Dit aspect mag niet over het hoofd worden gezien door recreatieve atleten. Net zoals de Stress Management Score, is de slaapscore van Fitbit een score tussen 1 en 100, met drie belangrijke componenten: duur, kwaliteit en herstel.

De component 'duur' houdt rekening met de totale slaapduur. Hoe langer de slaapduur, hoe beter. Echter, er wordt ook rekening gehouden met de tijd die iemand wakker ligt. De kwaliteit van de slaap wordt beoordeeld aan de hand van de tijd die wordt doorgebracht in de fasen van diepe slaap en REM-slaap. Voor sporters is met name diepe slaap van groot belang, omdat het lichaam zichzelf fysiek herstelt tijdens deze fase, wat essentieel is voor een optimaal herstel en voorbereiding op nieuwe trainingen. Het component 'herstel' biedt inzicht in hoe

”rustgevend”de slaap was. Ten eerste wordt gekeken naar hoe vaak iemand zich tijdens de slaap omdraait, aangezien dit kan duiden op een onrustige slaap. Vervolgens wordt ook de hartslag tijdens de slaap gemonitord. Normaal gesproken ligt de hartslag tijdens de slaap lager dan de rusthartslag. Wanneer de hartslag tijdens de slaap hoger is dan de rusthartslag, zal het algoritme een lagere score toekennen voor dit component. Hoewel de hartslag tijdens de slaap niet direct kan worden gecontroleerd, kunnen factoren zoals alcohol, ziekte en stress wel directe invloed uitoefenen op de hartslag tijdens de slaap.

Hartslag in rust Gemiddeld genomen ligt de normale hartslag in rust tussen de 50 en 90 slagen per minuut[82]. Deze wordt gemeten wanneer je tien minuten stilzit of direct na het wakker worden, terwijl je nog in bed ligt. Tijdens de slaap kan de hartslag nog lager zijn, wat een te verwachten fenomeen is. Dit wordt de Non-REM-hartslag genoemd en wordt niet beschouwd als de rusthartslag. Beide waarden zijn meestal terug te vinden in de applicatie van een smartwatch. Het is belangrijk op te merken dat de rusthartslag geen betrouwbare metriek is om direct te vergelijken met die van andere sporters. Een lagere rusthartslag betekent niet automatisch dat je fitter bent, aangezien de rusthartslag deels genetisch bepaald is. Bovendien hebben vrouwen doorgaans een iets hogere rusthartslag dan mannen[82]. Daarom is het belangrijk om het persoonlijke bereik van de rusthartslag in beschouwing te nemen om te bepalen of iemand al dan niet een hoge rusthartslag heeft.

Hartslag variabiliteit

Hartslagvariabiliteit (HRV) is een geheel andere parameter dan de gemiddelde hartslag. Naast de gemiddelde hartslag kan de hartslagvariabiliteit zowel van dag tot dag als van moment tot moment fluctueren. Een gemiddelde hartslag van bijvoorbeeld 60 slagen per minuut geeft alleen aan dat er gemiddeld 60 slagen per minuut plaatsvinden. Dit betekent echter niet dat er elke seconde een hartslag zal plaatsvinden. Tussen de hartslagen zit een kleine variatie, en dit is wat we hartslagvariabiliteit noemen. Deze variabiliteit wordt gemeten in milliseconden.

Bij de Fitbit Sense 2 wordt de hartslagvariabiliteit ’s nachts gedurende de diepe slaapfase berekend. Deze variabiliteit wordt gecontroleerd door het autonome zenuwstelsel, dat verantwoordelijk is voor meerdere processen in het lichaam, zoals ademhaling, vertering, hartslag en bloeddruk. Het autonome zenuwstelsel is verder verdeeld in twee grote delen: het sympathische en het parasympathische deel. Het parasympathische deel beïnvloedt organen die betrokken zijn bij rust en herstel in het lichaam, zoals het verteringsproces. Het sympathische deel daarentegen zet het lichaam in actie wanneer er sprake is van spanning of stress. Dit deel is verantwoordelijk voor de bekende ”fight or flight”-reactie van het lichaam. Wat interessant is voor sporters, is dat wanneer de hartslagvariabiliteit (HRV) daalt, het lichaam meer fysieke belasting ervaart. Dit kan erop wijzen dat het lichaam onvoldoende hersteld is van een training. Ook een opkomende ziekte, alcoholgebruik of een stressvolle periode op het werk kunnen een negatieve invloed hebben op de HRV. Belangrijk om op te merken is dat de HRV sterk afhankelijk is van persoon tot persoon. Zo kan een professionele atleet zelfs een lagere HRV hebben dan een gemiddelde recreatieve sporter, ondanks dat de eerste veel fitter is. Het is beter om te kijken naar het persoonlijke HRV-bereik van elke sporter om te bepalen of hij een hoge of lage HRV heeft op een gegeven dag.

4.1.2 Procedure van de datacollectie

De geselecteerde periode voor het dragen van de smartwatches was augustus, september en oktober, wat voor het onderzoek het meest geschikte moment was. Vanwege praktische beperkingen kon de dataverzameling niet vroeger starten tijdens het wielerseizoen, en ook niet in de wintermaanden omdat veel recreatieve wielrenners dan een winterstop houden of slechts korte indoor trainingen doen die mogelijk niet voldoende intensief zijn om het herstelproces goed te observeren. Voor zeer nauwkeurige resultaten met betrekking tot de voorspelling van het herstelproces van recreatieve wielrenners was een groter aantal deelnemers noodzakelijk.

Echter, vanwege budgettaire beperkingen was het niet mogelijk om meer dan vijf smartwatches beschikbaar te stellen. Een alternatieve benadering zou zijn geweest om proefpersonen de smartwatch slechts gedurende één maand te laten dragen, zodat dezelfde verzamelperiode driemaal herhaald kon worden met telkens vijf kandidaten, waardoor het aantal deelnemers uitgebreid had kunnen worden tot vijftien. Dit zou als voordeel hebben dat de dataset representatiever zou zijn voor een bredere doelgroep van recreatieve wielrenners dan wanneer slechts vijf deelnemers betrokken zouden zijn. In de praktijk is deze benadering echter minder eenvoudig te implementeren. De smartwatch vereist namelijk enkele dagen om zich aan de gebruiker aan te passen. Daarnaast trainen recreatieve wielrenners niet elke dag van de week, waardoor het aantal datapunten per deelnemer beperkt blijft. Om toch voldoende data per deelnemer te verzamelen, werd ervoor gekozen om het aantal deelnemers te beperken tot vijf, die gedurende drie maanden de smartwatch droegen.

Gedurende de datacollectie kreeg één deelnemer irritaties aan de pols door het bandje van de Fitbit. Om deze reden verzocht hij vroegtijdig te stoppen met het onderzoek. Aangezien de datacollectie al was gestart, kon er niet tijdig een andere deelnemer gevonden worden om de vrijgekomen smartwatch over te nemen. Omdat hij slechts enkele dagen de smartwatch heeft kunnen dragen, werd zijn data niet opgenomen in de dataset. Na de datacollectieperiode kwamen er nog problemen aan het licht met de data van twee andere deelnemers. Hoewel via de webinterface werd gecontroleerd of de deelnemers hun smartwatch regelmatig synchroniseerden met het Fitbit-platform, werden niet alle parameters correct gemeten of gesynchroniseerd. Dit probleem kan grotendeels worden toegeschreven aan de Fitbit-interface zelf. De webinterface verschilt aanzienlijk van de applicatie en op het moment van schrijven zijn bepaalde parameters slecht tot helemaal niet zichtbaar in de webinterface. Hierdoor werd ten onrechte aangenomen dat de datacollectie optimaal verliep. Bij één van die twee deelnemers, bleek dat het grootste deel van de belangrijke parameters niet werd gemeten of opgeslagen, waardoor de data van deze proefpersoon niet bruikbaar was. Van een andere proefpersoon werd de belangrijkste parameter, de Stress Management Score, niet correct geregistreerd. Ook deze data kon niet gebruikt worden. Uiteindelijk bleek de data van slechts twee deelnemers bruikbaar te zijn.

Twee deelnemers aan een dataset zijn niet representatief voor een gehele populatie. De onderzoeker zelf had toevallig ook een Fitbit smartwatch, zij het een ander model (Inspire 2), dat dezelfde belangrijke parameters kon registreren. Omdat de onderzoeker regelmatig fietst, werd zijn data ook geschikt geacht voor opname in de dataset, waardoor deze werd uitgebreid naar drie personen. Omdat de onderzoeker de Fitbit reeds lange tijd droeg, kon data over een volledig jaar worden verzameld, wat resulteerde in een groter aantal datapunten. Deze situatie leidde echter tot een ongelijke verdeling van data tussen de proefpersonen, wat de representativiteit van de dataset voor de gemiddelde recreatieve sporter mogelijk verminderde.

Drie deelnemers is nog steeds erg beperkt. Een oplossing werd gevonden bij een collega die ook de tool had ontwikkeld voor het omzetten van het Fitbit-archief in een leesbaar formaat. Hij had een studie uitgevoerd met vijf proefpersonen gedurende drie maanden om de invloed van alcohol op slaap te onderzoeken. Hoewel zijn studie geen specifieke criteria voor duursporters of wielrenners bevatte, werd besloten om zijn dataset op te nemen in de dataset voor deze masterproef, omdat dezelfde horloges werden gebruikt en de data met dezelfde tools kon worden verwerkt. Hierdoor kon zowel het aantal deelnemers als het aantal datapunten verder uitgebreid worden.

4.2 Verwerken van de data

Om deel te nemen aan de dataverzameling, werd aan de participanten gevraagd om een account aan te maken op het Fitbit-platform en hun e-mailadres te delen met de onderzoekers, zodat deze via de webinterface konden monitoren of alles goed verliep. Na deze periode werden alle smartwatches ingeleverd en werd de datacollectie stopgezet. In overeenstemming met de GDPR-wetgeving zijn platformen verplicht gebruikers inzage te geven in de opgeslagen gege-

vens. Hierdoor konden alle gegevens van de participanten eenvoudig worden opgevraagd via de webbrowserinterface van Fitbit. Vervolgens werd een zip-bestand met alle gegevens per participant verkregen. Via een collega werd een script verkregen om het zip-bestand naar een leesbaar formaat om te zetten.

Het script genereerde een dataset met de volgende gegevens, gedistilleerd uit het zip-bestand van Fitbit: datum, Stress Management Score, slaapscore, aantal stappen per dag, slaapduur, aantal minuten slaap in verschillende slaapfasen, aantal actieve minuten, aantal zittende minuten, overzicht van de sportactiviteiten op een dag, een overzicht van het aantal actieve minuten op elk moment van de dag (ochtend, middag, avond, nacht), een overzicht van de verschillende minuten in verschillende hartslagzones, hartslagvariabiliteit, NREMHR, ademhaling, gemiddelde zuurstofsaturatie.

De belangrijkste problemen met het script waren dat de rusthartslag niet automatisch werd verwerkt uit het originele zip-bestand. Gezien de relevantie van deze parameter, werd het script aangepast om deze functionaliteit te ondersteunen. Daarnaast verhoogde Fitbit het aantal gesportte minuten in bepaalde hartslagzone door gebruik te maken van een vermenigvuldigingsfactor. Omdat dit een sterk vertekend beeld geeft van de werkelijke activiteit van een atleet, werden de minuten gecorrigeerd naar de originele gesportte minuten.

Na analyse van de hartslaggegevens bleken sommige deelnemers een afwijkende gemiddelde hartslag te vertonen voor sommige sportactiviteiten. Uit een enquête onder deelnemers bleek dat de meerderheid een borstgedragen hartslagmeter gebruikte en de gegevens naar Strava uploadde. Significante verschillen werden vastgesteld bij het vergelijken van de gemiddelde hartslag tussen Strava en Fitbit. Omdat borstbanden als nauwkeuriger worden beschouwd dan smartwatches voor het meten van hartslag, wordt borstbanddata als de meest betrouwbare beschouwd in deze studie. Daarom werd een script ontwikkeld om activiteiten gemeten door de Fitbit te corrigeren met behulp van Strava-gegevens. Afwijkingen in hartslag hebben namelijk een sterke invloed op de schatting van de trainingsbelasting.

Een ander probleem met de nieuw gegenereerde dataset voor machine learning is de beperkte omvang van het aantal datapunten. In totaal konden ongeveer 800 datapunten worden verzameld gedurende drie maanden. Elk datapunt vertegenwoordigt precies één dag van een enkele persoon. Een dag omvat verschillende parameters die gemeten werden door de smartwatch. De belangrijkste gemeten parameters zijn terug te vinden in de bovenstaande sectie 4.1.1. Daarnaast worden de parameters van een bepaalde dag aangevuld met de belangrijkste gegevens van de volgende dag. Dit helpt bij het visualiseren van bepaalde relaties en bij het voorspellen van het herstel van de recreatieve atleet. Vanwege de beperkte dataset werd er gezocht naar alternatieven, maar online waren er weinig archieven met Fitbit-gegevens beschikbaar. Dit is niet verrassend, aangezien dergelijke (medische) gegevens meestal niet openbaar mogen worden gepubliceerd. Zo werd toch een onderzoek gevonden met Fitbit-gegevens van zestien personen, maar de dataset miste de essentiële Stress Management Scores en HRV-data, waarschijnlijk door het gebruik van een ouder smartwatchmodel. Vanwege het ontbreken van deze belangrijke gegevens werd besloten deze dataset niet op te nemen in de huidige studie.

4.3 Trainingsbelasting berekenen en verbeteren

Een parameter die niet door Fitbit werd gemeten of berekend, was de trainingsbelasting voor een bepaalde activiteit. Dit helpt bij het inschatten van de trainingsbelasting van een training. Zoals eerder vermeld, kan de werkelijke trainingsbelasting van de training voor het lichaam worden berekend op basis van de hartslag tijdens de training. Ook hier zijn verschillende methoden en modellen mogelijk zoals beschreven in sectie 2.8. Binnen deze masterproef is ervoor gekozen om te werken met een vijfzone-model voor de hartslag. Hoe hoger het zonenummer, des te meer trainingsbelasting deze zone genereert voor het lichaam. We gaan ervan uit dat één uur maximale inspanning een trainingsbelasting van 100 oplevert. Een gedetailleerde verdeling per zone is te vinden in de onderstaande tabel 4.1. Door deze nieuwe parameter te

introduceren, wordt het mogelijk om naast wielrennen ook andere sporten te ondersteunen. In deze masterproef wordt verondersteld dat voor elke sport dezelfde hartslagzones met dezelfde inspanningsniveaus gelden. Bovendien wordt hetzelfde hartslagmodel toegepast op elke recreatieve sporter die deelneemt en een smartwatch draagt. Het is echter mogelijk om de zones per sporter enigszins te personaliseren op basis van de maximale hartslag en/of in combinatie met de rusthartslag. Aangezien alle deelnemers jonge mannen waren met een leeftijd rond de twintig jaar, werd geoordeeld dat het hetzelfde vijfzone-model voor iedereen voldoende was voor de opzet van dit onderzoek.

Hartslagzone	Training belasting per uur
1 (laag)	20
1	30
1 (hoog)	40
2 (laag)	50
2 (hoog)	60
3	70
4	80
5a	100
5b	120
5c	140

Tabel 4.1: Trainingbelasting voor hartslagzones volgens TrainingPeaks ²

In plaats van alleen de trainingsbelasting per trainingssessie te berekenen, wordt in deze masterproef de totale belasting voor de hele dag berekend. Het is echter de bedoeling dat de recreatieve atleet het grootste deel van zijn belasting uit training haalt om fysiek beter te worden. In tegenstelling tot topsporters, die een zeer strikt schema volgen om optimaal te presteren tijdens wedstrijden, hebben de meeste recreatieve atleten dit niet. Recreatieve atleten kunnen naast hun training ook fysiek actief zijn met hun job, huishoudelijke taken, tuinieren, spelen met kinderen, enzovoort. Daarom is het belangrijk om deze extra fysieke activiteiten mee te nemen bij het berekenen van het herstel voor de volgende dag.

Hier zijn de concrete hartslagzones van de masterproef, gebaseerd op een jongeman van 25 jaar oud:

- Zone 1: Lager dan 117 slagen per minuut
- Zone 2: Tussen 118 en 147 slagen per minuut
- Zone 3: Tussen 148 en 156 slagen per minuut
- Zone 4: Tussen 157 en 176 slagen per minuut
- Zone 5: Meer dan 176 slagen per minuut

Met ook vervolgens trainingsbelasting per zone, afgeleid uit de bovenstaande tabel 4.1 van TrainingPeaks:

- Zone 1: 30 trainingsbelasting
- Zone 2: 55 trainingsbelasting
- Zone 3: 70 trainingsbelasting
- Zone 4: 80 trainingsbelasting
- Zone 5: 120 trainingsbelasting

Bij de toepassing van deze theorie met Fitbit smartwatches werd een probleem ontdekt. Fitbit biedt namelijk geen dagoverzicht van de hartslag aan dat is verdeeld in een vijfzone-model. In

plaats daarvan hanteert Fitbit een vierzone-model, waarbij de onderste zone eigenlijk een te lage hartslag weergeeft en daardoor niet correct is voor inspanning. De overige drie zones helpen de drager van het horloge in te schatten of hij in de vetverbrandingszone, de cardiozone of de maximale inspanningszone traint. Om dit vierzone-model om te zetten naar het vijfzone-model, werd de gemiddelde hartslag binnen de zones van Fitbit genomen. De hartslagverdeling voor het model van Fitbit was als volgt bij een van de deelnemers:

- Zone 1: Lager dan 107 slagen per minuut
- Zone 2: Tussen 108 en 136 slagen per minuut
- Zone 3: Tussen 137 en 173 slagen per minuut
- Zone 4: Meer dan 174 slagen per minuut

Gezien alle deelnemers mannelijk zijn en bijna dezelfde leeftijd hebben, zijn hun verschillende hartslagzones vergelijkbaar met het hierboven getoonde voorbeeld. Belangrijk om op te merken is dat Fitbit zelf het aantal minuten in elke hartslagzone voorziet in het zip-bestand. Hierdoor wordt voor elke atleet zijn individuele Fitbit-hartslagzonemodel toegepast, zonder dat de hartslagzones naar het bovenstaande voorbeeld hoeven te worden afgerond.

De hartslagzones zijn vervolgens vertaald naar de volgende trainingsbelasting per zone:

- Zone 1: 20 trainingsbelasting
- Zone 2: 35 trainingsbelasting
- Zone 3: 70 trainingsbelasting
- Zone 4: 120 trainingsbelasting

Dergelijke afrondingen zijn nooit zo nauwkeurig als het vijfzone-model, maar doordat ze consistent worden toegepast op alle data, kunnen ze toch een acceptabele schatting geven wanneer een vijfzone-model niet beschikbaar is.

De belasting op een dag wordt berekend door de minuten in verschillende Fitbit-zones te vermenigvuldigen met hun corresponderende belasting. Deze producten worden vervolgens opgeteld om de totale dagelijkse trainingsbelasting te bepalen. Uit het Fitbit-archief kan ook worden afgeleid of er fietsactiviteiten hebben plaatsgevonden, omdat deze automatisch worden geregistreerd wanneer de sporter de smartwatch draagt. Voor elke fietsactiviteit wordt de trainingsbelasting berekend op dezelfde manier als voor de gehele dag. Deze gegevens worden vergeleken met Strava om te controleren of de recreatieve sporter de activiteit ook op dit platform heeft geregistreerd, wat dient als extra controlemechanisme voor de nauwkeurigheid van de gegevens. Wanneer een activiteit op Strava is geregistreerd, is deze vaak nauwkeuriger gemeten met een fietscomputer en een hartslagmeter met borstband. Het koppelen van een Fitbit-activiteit aan een Strava-activiteit gebeurt op basis van het type activiteit (wielrennen) en de starttijden. Als de starttijden binnen een marge van tien minuten van elkaar liggen, wordt aangenomen dat het om dezelfde activiteit gaat. Omdat activiteiten van recreatieve wielrenners zelden korter dan tien minuten duren. Wanneer de fietsactiviteit zowel in het Fitbit- als Strava-archief voorkomt, wordt de voorkeur gegeven aan de nauwkeurigere Strava-gegevens. Vervolgens wordt de belasting voor de Strava-activiteit berekend met het vijfzone-hartslagmodel. Daarna wordt het verschil in trainingsbelasting tussen de Fitbit- en Strava-versies berekend en aangepast in de totale trainingsbelasting van de gehele dag. Analooq zal ook het de duratie van de fietsactiviteit aangepast worden naar de duratie die Strava vermeld heeft.

4.4 Analyseren, opschonen en verbeteren van de dataset

Na het verzamelen van de data en het berekenen van de trainingsbelasting, is het belangrijk om de dataset grondig te bestuderen met het oog op machine learning. Vanwege de beperkte omvang van de dataset wordt geprobeerd het aantal kenmerken te beperken. Het is namelijk eenvoudiger voor het machine learning-algoritme om te leren van een aantal beperkte, kwalitatieve kenmerken dan om te leren van een groot aantal kenmerken met veel onderlinge ruis.

Om te beginnen met de datacleaning, werd eerst berekend hoeveel kolommen datapunten misten en vervolgens hoeveel datapunten er in elke kolom ontbraken. In totaal waren er 22 verschillende kolommen waar datapunten ontbraken, waarbij het aantal ontbrekende datapunten per kolom varieerde. In het beste geval ontbrak slechts 1,4% van de gegevens, terwijl in het slechtste geval tot 95% van alle gegevens van de kolom ontbrak. Op basis van zowel de relevantie van de kolom als het percentage ontbrekende gegevens werd besloten of de kolom behouden of verwijderd moest worden. Acht kolommen werden verwijderd vanwege hun onbelangrijkheid voor toekomstige voorspellingen en het hoge percentage ontbrekende waarden (tussen 26% en 84%). Vier andere kolommen werden voorlopig behouden vanwege hun belangrijkheid bij het genereren van een nieuw kenmerk en het corrigeren van Fitbit-gegevens.

Vervolgens werden de parameters, waaronder de Stress Management Score (Fitbit), ook wel de readiness score genoemd in deze masterproef, evenals de hartslag in rust en hartslagvariabiliteit, aangepast naar een schaal van 1 tot 10. De originele score tussen 1 en 100 biedt te veel variatie in readiness scores, wat resulteert in slechts enkele datapunten per score. Door de scores te beperken tot een schaal van 1 tot 10, blijft er voldoende onderscheid tussen hoge en lage scores, terwijl het aantal mogelijke scores beperkt wordt, wat zorgt voor een betere data-analyse. De score werd herberekend door de readiness score van elke persoon afzonderlijk te bekijken en het bereik tussen de hoogste en laagste waarde in tien gelijke delen te verdelen. Afhankelijk van het compartiment waarin de originele readiness score viel, werd een nieuwe waarde toegewezen, oftewel het compartimentnummer. Op deze manier werd uniformiteit in de data van readiness scores gegarandeerd voor alle deelnemers in de dataset. Analoog aan deze herberekening werden ook de rusthartslag en hartslagvariabiliteit aangepast, omdat deze parameters sterk fluctueren en persoonsafhankelijk zijn, zoals besproken in sectie 4.1.1. Door deze uniformisatie wordt het mogelijk om de verschillende parameters te vergelijken tussen deelnemers, wat tevens waardevolle input vormt voor machine learning-toepassingen.

Na het uniformiseren van de bovenstaande parameters, werden de belangrijke parameters waarvoor er gegevens ontbraken verder aangevuld. Het betreft de volgende parameters: de tijd in de verschillende slaapfasen, de lengte van de slaap, de slaapscore, de hartslag in rust, de hartslagvariabiliteit en tot slot ook de readiness score. Vanwege hun belangrijkheid voor de voorspellingen en het zeer lage aantal ontbrekende waarden, is het de moeite waard om deze ontbrekende waarden aan te vullen. Maximaal ontbrak er slechts zes procent van alle gegevens van een bepaalde parameter. Er zijn verschillende strategieën mogelijk om deze waarden in te vullen, maar binnen deze masterproef is ervoor gekozen om de ontbrekende waarden te voorspellen met behulp van een Random Forest regressor. De rijen in de dataset waar alle waarden intact zijn, werden verzameld om een voorspelling te doen voor de rijen waar ten minste één waarde ontbrak. In de onderstaande tabel zijn alle R²-scores van de verschillende parameters te vinden waarvoor enkele ontbrekende waarden moesten worden voorspeld. De R²-score geeft een indicatie van in welke mate het model in staat is om de uitkomst te voorspellen. Deze score varieert tussen 0 en 1, waarbij 1 een perfect resultaat zou zijn.

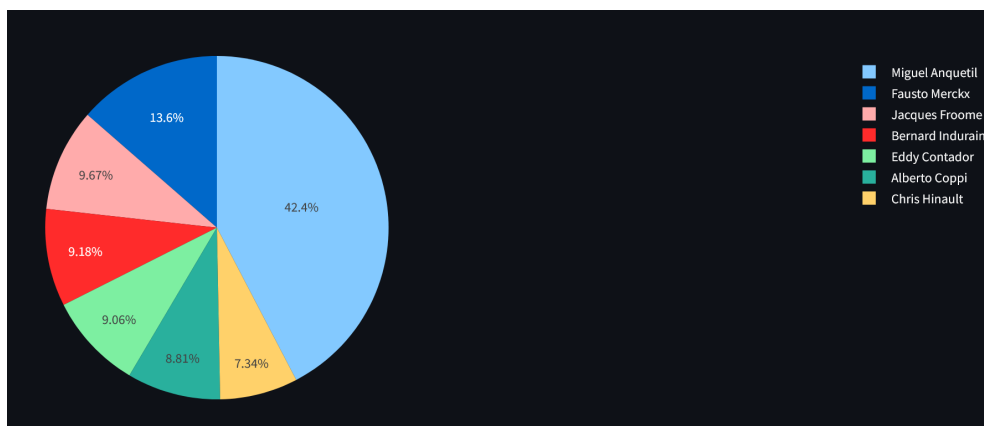
Parameter	R2 - score
Tijd wakker	0.965
Tijd in lichte slaap	0.990
Tijd in rem slaap	0.965
Tijd in diepe slaap	0.952
Slaap score	0.978
Totaal geslapen minuten	0.994
Hartslag in rust	0.870
Stress Management Score	0.904
Hartslag variabiliteit	0.919

Tabel 4.2: R2-score van de Random Forest Regressor op de belangrijke parameters waar datapunten ontbraken

Na al deze aanpassingen voor de dataset ontbreken er slechts waarden in twee kolommen: de trainingsbelasting per dag en het aantal fietsminuten. Deze ontbrekende waarden worden beschouwd als 0. In de praktijk komt deze vertaling overeen met het feit dat niet iedereen dagelijks fietst. Bovendien kunnen er dagen zijn waarop de drager van het horloge weinig fysieke activiteit heeft verricht.

4.5 Aanpassen dataset aan gekozen machine learning

Na het verbeteren en opschonen van de dataset, blijven er ongeveer 750 verschillende rijen over. Elke rij bevat de informatie van een volledige dag van een deelnemer. De verdeling per deelnemer is te zien in het onderstaande diagram. Let op, de echte namen van de personen zijn vervangen door pseudoniemen om hun identiteit te beschermen. Op onderstaande figuur (zie 4.1) is de verdeling van datapunten tussen de deelnemers zichtbaar. Uit de figuur is op te merken dat een groot deel van de datapunten slechts van een deelnemer zijn.

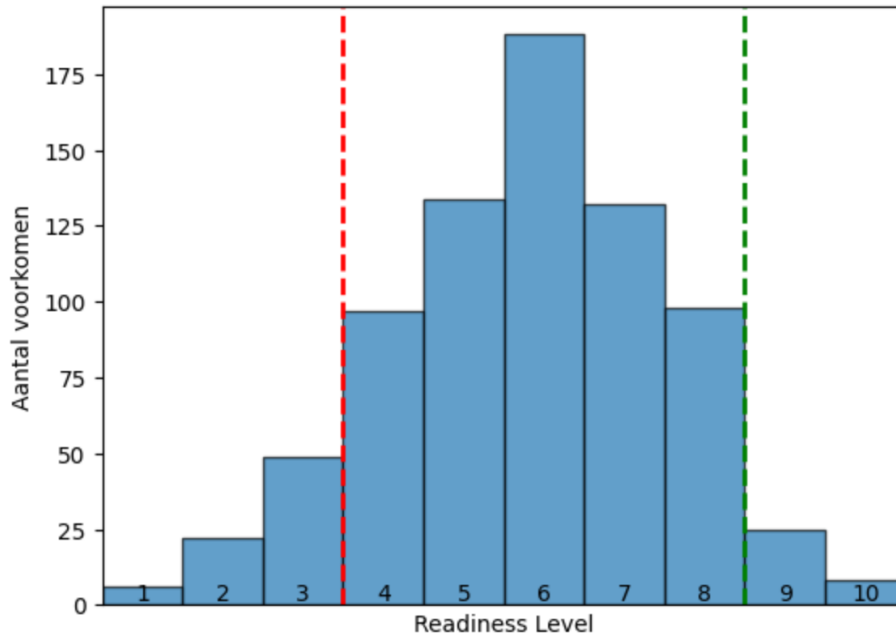


Figuur 4.1: Verdeling verschillende deelnemers in dataset

De zojuist gemaakte dataset zal vervolgens als input gebruikt worden voor machine learning. De bedoeling is om op basis van alle gegevens van een gegeven dag, een inschatting te maken van het readiness score van de atleet voor de volgende dag. Waarmee het readiness score een indicatie moet geven over de trainingsparaatheid van de recreatieve atleet. Op basis van de kennis van de dataset zal een machine learning algoritme gekozen worden.

Een gedetailleerde verdeling van de readiness score tussen 1 en 10 is weergegeven in figuur 4.2, die een patroon vertoont dat lijkt op een normale verdeling. Om het scenario van matig herstel bij recreatieve sporters te simuleren, zijn de waarden 1, 2 en 3 gemapt naar het label 1. Dit

komt overeen met de waarden voor de rode lijn in figuur 4.2. Waarden 4 tot en met 8 zijn gemapt naar het label 2, wat overeenkomt met een gemiddeld herstel. Dit wordt gerepresenteerd door waarden tussen de rode en de groene lijn in figuur 4.2. De overige waarden, 9 en 10, zijn gemapt naar het label 3, wat aangeeft dat er sprake is van bovengemiddeld herstel. De grenzen voor deze nieuwe labels zijn enigszins arbitrair gekozen, zonder een strikt wetenschappelijke rechtvaardiging voor de exacte locatie. In dit onderzoek zijn ze zorgvuldig gekozen om een relatief gelijkmatige verdeling van slechte en bovengemiddelde herstelniveaus te verkrijgen, waarbij beide nieuwe klassen ongeveer gelijke grootte hebben. Dit weerspiegelt de realiteit, waarin de meeste dagen als gemiddeld zouden moeten worden ervaren, met slechts een beperkt aantal dagen die een slecht of uitstekend herstel laten zien.



Figuur 4.2: Verdeling readiness score

Binnen het domein van machine learning zijn er diverse methodologische benaderingen beschikbaar, zoals supervised learning, unsupervised learning, reinforced learning, of een combinatie daarvan. Om een antwoord te formuleren op de gestelde probleemstelling, werd eerst teruggegrepen naar de oorspronkelijke vraag: het voorspellen van het herstel van de sporter één dag later, met gelijktijdig inzicht in de samenstellende factoren van dit herstel. Vanuit de dataset is een "readiness score" beschikbaar dat kan dienen als label om het herstel te beoordelen. Alle andere kenmerken hebben duidelijk gelabelde waarden, inclusief een "readiness score" voor zowel de huidige dag als de dag erna. Alle benodigde inputgegevens zijn beschikbaar, evenals de verwachte uitkomst. Supervised learning vormt de meest geschikte methodologie om met gelabelde invoer- en uitvoergegevens te zoeken naar een verband tussen beiden.

Vervolgens moet er binnen de tak van supervised learning een keuze worden gemaakt tussen classificatie en regressie. Met regressie zou in een ideaal scenario perfect kunnen worden voorspeld hoe goed een recreatieve sporter morgen zal herstellen op basis van de gemeten parameters van vandaag. Echter, gezien de dataset klein is en mogelijk vertekend in het voordeel van één van de proefpersonen, blijkt dit perfecte scenario moeilijk te realiseren. Om de nauwkeurigheid van het model te verbeteren, wordt de doelstelling van het model aangepast. Het herstel van de recreatieve sporter zal worden onderverdeeld in drie belangrijke categorieën: onvoldoende hersteld, voldoende hersteld, en bovengemiddeld hersteld. Om deze classificatie mogelijk te maken, werden de labels van de readiness score in de dataset herschaald naar labels tussen 1 en 3. Label 1 duidt op ontoereikend herstel van de sporter, terwijl label 3 aangeeft dat de

sporter bovengemiddeld hersteld is. Binnen deze implementatie geven de strikte grenzen tussen onvoldoende herstel en voldoende herstel echter een duidelijke scheidslijn aan. Dit is een beperking binnen de masterproef. Het zou wellicht beter zijn geweest om labels te ontwerpen zoals: onvoldoende herstel, grijze zone en voldoende herstel. Op deze manier kon duidelijker naar recreatieve atleten worden gecommuniceerd dat herstel niet zo zwart-wit is, maar een combinatie van veel factoren.

4.6 Test procedure voor machine learning resultaten

Om consistent te zijn met alle performantiewaarden die in het volgende deel van de masterproef vermeld zullen worden, werden de volgende maatregelen getroffen omtrent het testen van verschillende machine learning modellen met de dataset:

- Training en test set werden consistent verdeeld in 75 / 25 over alle testen heen. Daarnaast werd altijd dezelfde *random state* optie gebruikt in combinatie met de stratify optie.
- Aan elk machine learning model werd dezelfde *random state* toegevoegd. Zo had ieder model dezelfde random state en konden reproduceerbare resultaten verkregen worden die vervolgens te vergelijken zijn met andere testen in de masterproef.
- Gebruik maken van kruisvalidatie

Deze opties vergemakkelijken de vergelijking van verschillende modellen. De prestaties van het machine learning-algoritme worden in sterke mate beïnvloed door zowel de training- als testdatasets, met name wanneer er sprake is van een onevenwichtige verdeling van klassen, zoals in deze masterproef. De verdeling van de klassen is als volgt: 10 / 85 / 5. Door gebruik te maken van de stratify-optie wordt ervoor gezorgd dat zowel in de test- als trainingsset dezelfde verhouding van datapunten voorkomt als in de oorspronkelijke dataset. Wat de meest eerlijke manier is om te trainen en te berekenen wat de performantie van elk model is op elke klasse van de dataset. Voor verdere uniformiteit wordt de *random state*-optie gebruikt, waardoor het machine learning-algoritme op een consistente wijze probeert te leren. Het is echter geen garantie dat exact hetzelfde wordt geleerd bij elke uitvoering, maar het biedt de beste benadering mogelijk, wat belangrijk is voor consistentie over verschillende tests.

Tot slot wordt ook kruisvalidatie toegepast. Met deze methode wordt eenzelfde model herhaaldelijk getest, waarna het gemiddelde van de prestaties wordt genomen. Het is belangrijk op te merken dat elk datapunt zowel eens in de training als in de testdataset kan voorkomen bij deze aanpak. Hierdoor worden alle beschikbare gegevens maximaal benut tijdens het testen van een model. Aangezien deze masterproef beschikt over een kleine dataset, is het wenselijk om het gebruik van elk datapunt te maximaliseren.

Hoofdstuk 5

Resultaten

Voor deze masterproef werden vier verschillende algoritmen uitgetest, omdat het vooraf moeilijk te bepalen is welke dataset het best scoort. Zo werden: GradientBoosting, Adaboost, Decision Trees en Random Forest met elkaar vergeleken. Alle testen van de verschillende modellen werden onderworpen aan een uniforme testprocedure, waarbij gebruik werd gemaakt van kruisvalidatie. Hierbij dient echter te worden opgemerkt dat de kruisvalidatie, zoals beschreven in sectie 3.4, werd aangepast van het standaard k-fold algoritme naar het gestratificeerde k-fold algoritme. Vanwege de kleine dataset en de onbalans in de klassenverhoudingen is het namelijk belangrijk om de dataset niet willekeurig in k delen te splitsen, maar de k verschillende delen zodanig op te splitsen dat alle delen ongeveer dezelfde klassenverhoudingen hebben als de oorspronkelijke dataset.

5.1 Resultaten zonder optimalisaties

	Gradient Boosting	Random Forest	Adaboost	Decision Trees
Accuraatheid	83.1%	85.2%	70.1%	73.8%
Gewogen F1	78.5%	78.9%	72.7%	74.5%
F1	33.9%	31.5%	41.7%	36.4%
Gewogen Precisie	75.5%	75.1%	76.6%	75.3%
Precisie	37.4%	35.2%	41.2%	36.5%
Gewogen Sensitiviteit	83.1%	85.2%	70.1%	73.8%
Sensitiviteit	34.6%	33.6%	44.7%	36.4%

Tabel 5.1: Resultaten zonder optimalisaties

In de bovenstaande tabel lijken het Gradient Boosting en het Random Forest model duidelijk de betere modellen met een zeer goede accuraatheid. Echter, het is belangrijk om te benadrukken dat er drie aparte klassen in de dataset zijn met een zeer ongelijke verdeling. De eerste klasse bevat ongeveer tien procent van alle datapunten in de dataset, de tweede klasse ongeveer 85 procent, en de derde klasse slechts vijf procent. Vanwege deze klasse-onbalans is accuraatheid niet altijd de beste manier om de prestaties van een model te evalueren. Als voorbeeld, stel dat een model zoals Random Forest 100 keer voorspelt dat klasse 2 de correcte oplossing is. Als we aannemen dat de werkelijke verdeling van de antwoorden dezelfde is als die van de dataset (10/85/5), zou het model nog steeds een accuraatheid van 85 procent behalen. Een belangrijke kanttekening bij deze resultaten is dat het model mogelijk niet in staat is om klasse 1 en 3 effectief te voorspellen. In een dergelijke situatie zou het eenvoudig zijn om elke voorspelling automatisch als klasse 2 te labelen, zonder zelfs een machine learning model te hoeven trainen.

Daarom is het essentieel om ook andere evaluatiemetrieken van het model te bekijken, zoals de sensitiviteit en precisie.

Bij het evalueren van de gewogen metrieken lijken de scores relatief goed. Dit komt doordat de metrieken gewogen zijn op basis van de grootte van de klassen, waardoor de omvang van klasse 2 een grote invloed heeft op de uiteindelijke scores. Wanneer echter naar de algemene, ongewogen metrieken wordt gekeken, blijkt dat deze slechts de helft van de gewogen metrieken bedragen. Dit significant lagere resultaat kan worden toegeschreven aan het feit dat de voorspellingen voor de grootste klasse (klasse 2) uitstekend presteren, terwijl de prestaties voor de twee kleinere klassen (klasse 1 en 3) achterblijven. Neem een extreem, maar realistisch scenario waarin de voorspelling voor de grootste klasse goed werkt. Stel dat de precisie voor de grootste klasse 90 procent bedraagt, terwijl de precisie voor de twee kleinere klassen in beide gevallen 0 procent is. Beschouw vervolgens de verdeling van de dataset (10/85/5) over de verschillende klassen. De gewogen precisie bedraagt dan 76,5 procent, berekend als $0 \times 0.10 + 0.90 \times 0.85 + 0 \times 0.05$. Dit lijkt een relatief goede score voor een model. Echter, wanneer de algemene precisie wordt berekend, zal het model slechts 30 procent scoren, aangezien $(0+0.90+0)/3$. Hier worden de precisiepercentages van alle klassen opgeteld en gedeeld door het aantal klassen om de uiteindelijke score te verkrijgen. Een soortgelijke berekening geldt ook voor andere metrieken.

Om te bepalen welk model het beste presteert, zullen de volgende metrieken tegen elkaar worden afgewogen: accuraatheid, (gewogen) precisie en (gewogen) sensitiviteit, zie sectie 3.3 voor meer informatie rond de metrieken. De keuze voor accuraatheid is relatief vanzelfsprekend. Ondanks de ongebalanceerde dataset, blijft het wenselijk dat een algoritme een hoge mate van accuraatheid behaalt. Deze eenvoudige metriek wordt vaak gebruikt in de praktijk en is daardoor herkenbaar en begrijpelijk voor alle gebruikers. Naast accuraatheid zal ook precisie in overweging worden genomen bij het evalueren van de modellen. Wanneer de precisie groot is, zal een atleet waarschijnlijk eerder geneigd zijn om de aanbeveling van het model op te volgen. Echter, wanneer het model voorspelt dat een atleet voldoende hersteld is terwijl dit niet het geval is, kan dit negatieve gevolgen hebben, zoals overtraining. Dit benadrukt het belang van een hoge precisie om dergelijke ongunstige uitkomsten te voorkomen. Tot slot zal ook de sensitiviteit in overweging worden genomen, hoewel de focus voornamelijk ligt op accuraatheid en precisie. Op deze manier kan een algemeen goed presterend algoritme worden gewaarborgd.

Bij de toepassing van de criteria uit de vorige paragraaf wordt duidelijk dat één algoritme aanzienlijk lagere resultaten behaalt dan de andere drie. Het Decision Tree-model presteert in vergelijking met Gradient Boosting en Random Forest aanzienlijk minder op het gebied van accuraatheid. Daarom gaat de voorkeur uit naar de laatstgenoemde twee algoritmen. Ondanks dat het Decision Tree-model een iets hogere accuraatheid behaalde dan AdaBoost, is de precisie aanzienlijk lager, waardoor Decision Trees niet als de beste keuze kan worden beschouwd.

Het beoordelen van welk van de overgebleven drie modellen het beste is, blijkt een uitdaging. Gradient Boosting en Random Forest presteren zo vergelijkbaar dat de verschillen momenteel verwaarloosbaar zijn. AdaBoost daarentegen vertoont opmerkelijk andere prestaties. Hier kan een afweging gemaakt worden tussen accuraatheid, precisie en sensitiviteit. De keuze is dus niet zo eenduidig; deze is sterk afhankelijk van de parameters waar de meeste nadruk op wordt gelegd.

5.2 Geoptimaliseerde *hyperparameters*

Uit de vergelijking van de overeenkomstige gewogen en ongewogen parameters van alle kandidaatmodellen blijkt duidelijk dat de modellen goede prestaties leveren op de meerderheidsklasse, maar teleurstellend presteren op de minderheidsklasse. Een essentiële stap in het optimaliseren van het model voor een specifieke dataset is het verfijnen van de *hyperparameters* ervan. Sommige combinaties van *hyperparameters* kunnen effectiever zijn dan andere voor de gegeven dataset.

Het voorspellen van een goede combinatie van optimale *hyperparameters* blijft echter een uitdaging. Het verfijnen van *hyperparameters* is een arbeidsintensief proces waarbij systematisch elke combinatie van parameters moet worden getest en de resultaten nauwkeurig moeten worden geëvalueerd. Na afloop kan, door middel van grondige analyses met de nadruk op één van de prestatiecriteria, de meest geschikte combinatie van *hyperparameters* worden gekozen. Bovendien verschillen de mogelijke *hyperparameters* voor elk model, aangezien elk model verschillende eigenschappen heeft die moeten worden afgestemd tijdens het trainen en het maken van nauwkeurige voorspellingen.

Als eerste worden de *hyperparameters* van Gradient Boosting en Adaboost besproken. Beide modellen maken gebruik van de boosting-techniek en delen daarom ook een aantal dezelfde *hyperparameters*. De volgende parameters zijn relevant voor deze boosting-modellen:

- *N_estimators*: Het maximale aantal schatters waarbij de boosting zal worden gestopt. In het geval van een ideale fit, zal de de training vroegtijdig gestopt worden.
- *Learning_rate*: Dit is een gewicht dat toegepast wordt bij de classifier bij elke iteratie van de boosting. Er moet een afweging gemaakt worden tussen de leersnelheid en het aantal schatters.
- Maximum diepte: De maximum diepte van elke individuele schatter. De maximale diepte beperkt het aantal knooppunten in de boom. Het is mogelijk deze parameter te beperken zodat er een maximum diepte gezet is, maar het is niet verplicht. (enkel voor Gradient Boosting)

Bovenstaande lijst omvat niet alle verschillende *hyperparameters* die aangepast kunnen worden, er zijn er nog verschillende meer voor beide modellen.

Naast de modellen die worden ondersteund door boosting, zijn er ook nog de modellen Random Forest en Decision Trees die voornamelijk steunen op het gebruik van bomen. Voor het Decision Tree-model werd gekozen om de volgende *hyperparameters* te optimaliseren:

- *Min_samples_split*: Het minimaal aantal samples dat nodig is om een intern knooppunt verder op te splitsen.
- Maximum diepte: De maximale diepte van de boom. Deze mag uitzetten totdat alle bladeren puur zijn of totdat de bladen minder samples bevatten dan de *Min_samples_split*.

Bij Random Forest werden daarentegen meerdere *hyperparameters* geselecteerd voor optimalisatie. Dit besluit werd genomen op basis van bestaande kennis waaruit blijkt dat Random Forest in de praktijk bij andere toepassingen ook in staat is om goede resultaten te behalen. De lijst met te optimaliseren *hyperparameters* voor Random Forest is als volgt:

- *N_estimators*: Het aantal beslissingsbomen in het totale bos. Meestal zorgt een hoger aantal beslissingsbomen voor een verbeterde performantie. Daartegenover staat dan dat de computationele kost voor training en predicties zal toenemen.
- *Min_samples_split*: Het minimum aantal samples dat nodig is om een intern knooppunt verder te splitsen.
- *Min_samples_leaf*: Het minimum aantal data samples dat vereist is om zich in een bladknooppunt te bevinden. Een splitsingspunt wordt alleen overwogen als er na de splitsing minimaal *class_weight* aantal samples aanwezig is in zowel de linker- als rechterside. Het minimum aantal samples kan gedefinieerd worden door de *Min_samples_leaf* parameter aan te passen.
- *Class_weight*: Gewichten geassocieerd met de klassen. Indien niet opgegeven, wordt verondersteld dat alle klassen een gewicht van één hebben. Ook bestaat er de optie "gebalanceerd" waarbij de gewichten omgekeerd evenredig aangepast worden aan de klassenfrequentie. (Formule: $n_samples / (n_classes * np.bincount(y))$). Dit helpt om minder voorkomende klassen een groter gewicht te geven bij het trainen.

Zowel de lijst van *hyperparameters* voor het Decision Tree-model als voor het Random Forest-model is niet eindig. Er zijn nog verschillende andere parameters die verder verfijnd kunnen worden voor elk van deze modellen. Vanwege tijdsbeperkingen was het echter niet haalbaar om alle mogelijke parameters voor elk model te testen. Voor elk model zijn slechts een aantal geselecteerde parameters geoptimaliseerd. Zelfs bij deze selectie van parameters is het onmogelijk om alle mogelijke combinaties te testen. Bijvoorbeeld, alleen al het bereik van de 'maximum diepte' is een continu interval tussen 1 en oneindig.

Om een voldoende aantal combinaties van *hyperparameters* in de praktijk te kunnen testen, werd in eerste instantie gekeken naar de standaardwaarden van de te optimaliseren *hyperparameters*. Op basis hiervan werd een inschatting gemaakt van de waarden die realistisch zouden zijn om te testen, om zo tijd te besparen en te voorkomen dat onrealistische waarden werden getest. Vervolgens werden voor elke *hyperparameter* verschillende mogelijke waarden gekozen en opgeslagen. Alle combinaties van deze verschillende *hyperparameters* werden vervolgens geëvalueerd met behulp van een *gridsearch*-implementatie. Tijdens het zoeken naar de optimale parameters met behulp van *gridsearch* moest ook worden bepaald welke metriek als de belangrijkste werd beschouwd. Momenteel ondersteunt de implementatie van *gridsearch* niet om meerdere metrieken tegelijk te optimaliseren.

Voor deze masterproef werd gekozen om de precisie te optimaliseren. Uit eerdere resultaten bleek namelijk dat de accuraatheid al zeer goed was, waardoor daar geen verdere optimalisatie nodig was. Echter, de precisie en de sensitiviteit waren niet op hetzelfde niveau als de accuraatheid. De keuze voor precisie boven sensitiviteit werd gemaakt vanwege het belang dat werd gehecht aan de kwaliteit van de positieve voorspellingen. In de onderstaande lijst zijn de optimale *hyperparameters* voor elk model terug te vinden.

- Gradient Boosting
 - *N_estimators*: 55
 - *Learning_rate*: 0.1
 - Maximum diepte: 3
- Random Forest
 - *N_estimators*: 30
 - *Min_samples_split*: 2
 - *Min_samples_leaf*: 4
 - *Class_weight*: gebalanceerd
- Adaboost
 - *N_estimators*: 35
 - *Learning_rate*: 1.25
- Decision Trees
 - *Min_samples_split*: 3
 - Maximum diepte: *None* (geen begrenzing)

	Gradient Boosting	Random Forest	Adaboost	Decision Trees
Accuraatheid	83.9%	82.6%	67.5%	73.7%
Gewogen F1	78.7%	79.0%	71.1%	74.7%
F1	33.6%	39.2%	43.4%	37.4%
Gewogen Precisie	76.6%	77.5%	77.9%	76.0%
Precisie	41.9%	44.9%	43.4%	37.0%
Gewogen Sensitiviteit	83.9%	82.6%	67.5%	73.7%
Sensitiviteit	34.5%	39.0%	49.7%	38.2%

Tabel 5.2: Resultaten met geoptimaliseerde *hyperparameters*

De resultaten tonen aan dat de gewogen precisie bijna identiek is aan die van de modellen zonder geoptimaliseerde *hyperparameters*. Hoewel de cijfers in de meeste gevallen een lichte verbetering lijken te vertonen, blijven de verbeteringen zo minimaal dat het niet met absolute zekerheid kan worden gezegd of ze significant zijn, of dat ze binnen de foutmarge van het testen vallen. Niettemin valt op dat de ongewogen precisie aanzienlijk is verbeterd bij zowel het Gradient Boosting- als het Random Forest-model. Bij Adaboost lijkt een kleine verbetering te zijn bereikt, maar bij het Decision Tree-model lijkt er geen verbetering mogelijk te zijn.

Het is ook opmerkelijk dat de eerder behaalde goede accuraatheid niet significant is verminderd door het optimaliseren van het model voor een andere metriek. Hoewel er een kleine afname van de accuraatheid kan worden waargenomen bij het Adaboost- en Random Forest-model, blijft deze afname in verhouding tot de extra precisie een acceptabel compromis tussen beide metrieken.

5.3 SMOTE en geoptimaliseerde *hyperparameters*

Zoals eerder vermeld, zijn er methoden om de prestaties van ongebalanceerde datasets te verbeteren, en een van deze methoden is de SMOTE-techniek. In deze masterproef is besloten om beide minderheidsklassen te oversamplen, met als doel betere resultaten te behalen. De oorspronkelijke verdeling van de drie klassen in de dataset was 10% / 85% / 5%.

Om de optimale verhouding tussen echte en gesynthetiseerde data te vinden, zijn verschillende configuraties getest. Hierbij is geen rekening gehouden met de verhouding tussen de twee minderheidsklassen; beide zijn met behulp van SMOTE op gelijk niveau gebracht. Deze configuraties zijn niet meer uitgedrukt in percentages, maar in een verhouding tussen beide delen. Neem bijvoorbeeld "1 / 8 / 1", wat betekent dat voor elke 8 *samples* van de meerderheidsklasse exact één *sample* is van elke minderheidsklasse (klasse 1 en klasse 3). De volgende configuraties zijn allemaal getest:

- 1 / 8 / 1
- 2 / 8 / 2
- 3 / 8 / 3
- 4 / 8 / 4
- 5 / 8 / 5
- 6 / 8 / 6
- 7 / 8 / 7
- 8 / 8 / 8

Het is essentieel op te merken dat het testen van verschillende verhoudingen tussen echte en gesynthetiseerde data op een analoge manier is verlopen als de optimalisatie van enkel de

hyperparameters. Bovendien zijn tegelijkertijd met de testen van de ideale verhouding tussen echte en gesynthetiseerde data ook de ideale *hyperparameters* opnieuw getest. Deze stap is noodzakelijk omdat de dataset opnieuw is aangepast en er gestreefd wordt naar het creëren van een beter model dan de vorige iteratie.

In onderstaande tabel is de zijn de beste *hyperparameters* en de ideale SMOTE instellingen per model terug te vinden.

- Gradient Boosting
 - *N_estimators*: 30
 - *Learning_rate*: 0.08
 - Maximum diepte: 1
 - SMOTE-verdeling: 3 / 8 / 3
- Random Forest
 - *N_estimators*: 25
 - *Min_samples_split*: 2
 - *Min_samples_leaf*: 1
 - *Class_weight*: None (geen extra balans)
 - SMOTE-verdeling: 1 / 8 / 1
- Adaboost
 - *N_estimators*: 30
 - *Learning_rate*: 1.25
 - SMOTE-verdeling: 7 / 8 / 7
- Decision Trees
 - *Min_samples_split*: 0.1
 - Maximum diepte: 5
 - SMOTE-verdeling: 8 / 8 / 8

In de volgende onderstaande tabel zijn de resultaten terug te vinden die de modellen met bovenstaande parameters hebben kunnen produceren.

	Gradient Boosting	Random Forest	Adaboost	Decision Trees
Accuraatheid	82.5%	82.6%	70.0%	73.0%
Gewogen F1	79.0%	79.4%	72.4%	74.3%
F1	34.1%	41.8%	41.5%	36.9%
Gewogen Precisie	76.7%	77.1%	76.2%	75.9%
Precisie	41.6%	44.4%	40.4%	37.1%
Gewogen Sensitiviteit	84.1%	82.6%	70.0%	73.0%
Sensitiviteit	34.8%	41.6%	46.6%	37.2%

Tabel 5.3: Resultaten met SMOTE en geoptimaliseerde *hyperparameters*

Uit de tabel kan worden afgeleid dat sommige metrieken een licht hogere score lijken te hebben door het gebruik van SMOTE, terwijl andere metrieken een licht lagere score lijken te hebben dan wanneer alleen de *hyperparameters* werden geoptimaliseerd. Met andere woorden, de verschillen tussen deze resultaten en de vorige zijn verwaarloosbaar. Het lijkt er dus op

dat verdere optimalisatie door middel van SMOTE op de dataset niet het gewenste resultaat oplevert.

Uiteindelijk moet een model worden gekozen dat vervolgens geïmplementeerd kan worden in een website of app waarmee gebruikers aan de slag kunnen gaan bij deze masterproef. Vanwege consistente en betere prestaties wordt gekozen voor het Random Forest model. Dit model heeft een zeer hoge accuraatheid en de precisie is zowel bij deze resultaten als bij de vorige resultaten de hoogste van alle geteste modellen.

Hoofdstuk 6

Webapplicatie: voorspellen van herstelstatus recreatieve atleet

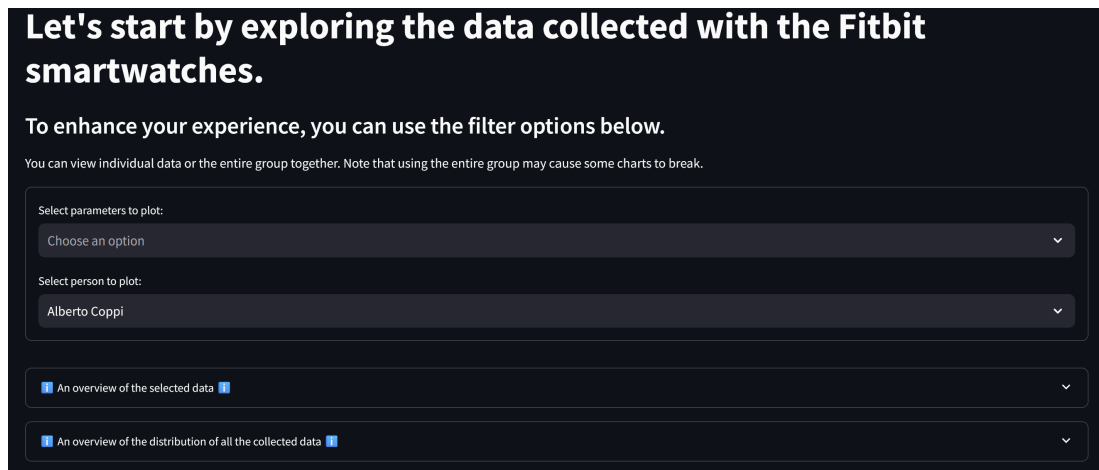
Om de deelnemers van het onderzoek meer inzicht te geven in de voorspellingen van het ontwikkelde model, werd een webapplicatie ontworpen. Deze applicatie is ontwikkeld met behulp van het Python-framework Streamlit¹. Streamlit is specifiek ontworpen voor toepassingen in datawetenschap en machine learning en biedt standaard ondersteuning voor het weergeven van Pandas DataFrames en diverse visualisaties met Seaborn en Matplotlib. Dankzij dit framework is er geen behoefte aan een complexe frontend of backend. Beide kunnen eenvoudig vanuit Python worden aangestuurd zonder dat er uitgebreide kennis van webontwikkeling vereist is.

6.1 Verkennen van de dataset en eenvoudige relaties

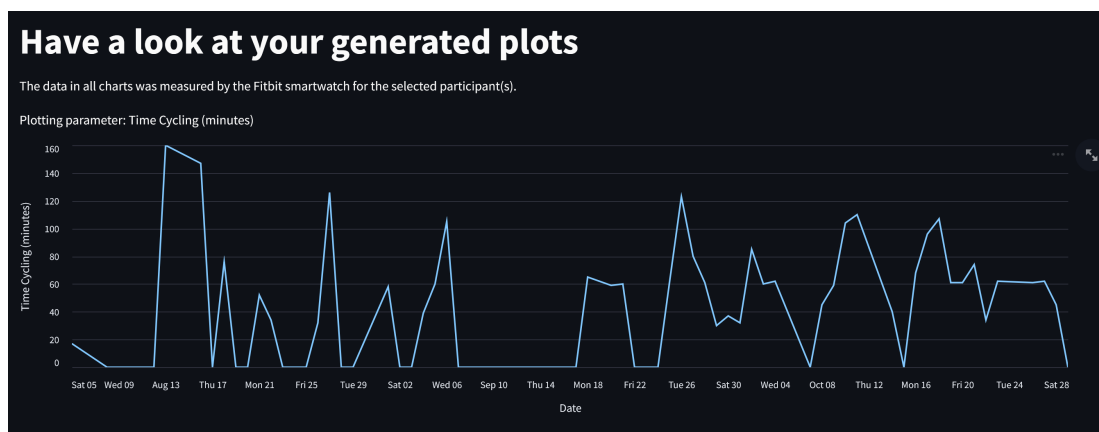
De eerste sectie in de applicatie beschrijft het doel van de thesis, zodat gebruikers een duidelijk beeld krijgen van de inhoud en het doel van de applicatie. Vervolgens is er een sectie waarin gebruikers de dataset kunnen verkennen, zoals te zien is in figuur 6.1. Deze visualisatie biedt een overzicht van alle datapunten in de dataset en toont tevens het aandeel van elke deelnemer in de dataset. Echter, alleen de datapunten weergeven maakt het lastig voor gebruikers om relaties tussen verschillende parameters te ontdekken. Daarom biedt de applicatie de mogelijkheid om een specifieke persoon uit de dataset te selecteren en vervolgens diverse interessante parameters te bekijken. Elke parameter wordt over de tijd weergegeven, waardoor eventuele uitschieters zichtbaar worden. Niet alle parameters uit de dataset zijn beschikbaar voor inspectie in deze visualisatie. Om verwarring bij de gebruikers te voorkomen, zijn alleen de meest relevante parameters met betrekking tot duursporten geselecteerd. Deze parameters omvatten: het aantal stappen per dag, de hartslag in rust, de hartslagvariabiliteit, de *heart rate load* (trainingsbelasting over geheel de dag) en de tijd gefietst per dag.

In de hierop volgende visualisatie is een grafiek te zien die is gegenereerd op basis van de voorkeuren van de gebruiker. Dit kan gezien worden in figuur 6.2 Het is een bewuste keuze om eerst de voorkeuren van de gebruiker te vragen voordat grafieken worden getoond. Hierdoor ziet de gebruiker alleen de grafieken die relevant zijn voor zijn interessegebied, wat verwarring over verschillende grafieken van verschillende deelnemers in de dataset voorkomt. Verder wordt in deze specifieke grafiek weergegeven hoe vaak en hoe lang een deelnemer heeft gefietst.

¹<https://streamlit.io/>



Figuur 6.1: Kiezen van gewenste grafieken en personen

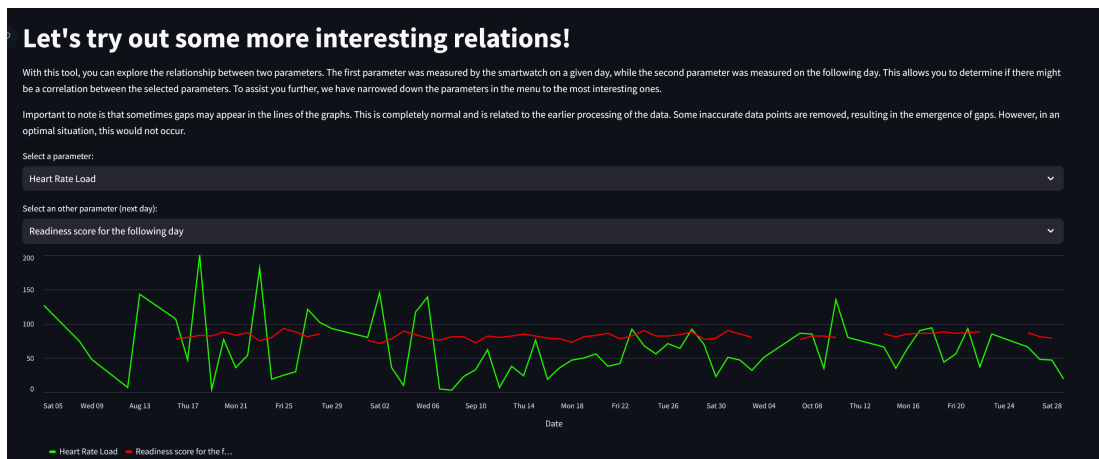


Figuur 6.2: Voorbeeld van een gegenereerde grafiek

6.2 Verkennen van complexe relaties

In de vorige visualisaties was het slechts mogelijk om één parameter over de tijd te inspecteren. Echter, het kan ook interessant zijn om twee verschillende parameters in dezelfde grafiek te plotten om te analyseren of er een direct verband tussen deze parameters bestaat. Ook deze grafieken moeten door de gebruiker zelf gegenereerd worden. Standaard worden niet alle mogelijke grafieken getoond om het overzicht binnen de applicatie te behouden. Daarnaast is de lijst met mogelijke opties om te inspecteren beperkt tot de meest interessante relaties. Deze visualisatie 6.3 probeert de invloed van een parameter, gemeten op een bepaalde dag, op een andere parameter, gemeten exact een dag later, te visualiseren.

In een ideaal scenario zou bijvoorbeeld zichtbaar zijn dat een zeer zware fietstocht van gisteren de readiness score van vandaag verlaagt, omdat het lichaam onvoldoende hersteld is. De parameters die voor een gegeven dag geselecteerd kunnen worden, zijn beperkt tot: *heart rate load* (trainingsbelasting over geheel de dag) en totale fietsminuten. De parameters voor de dag erna zijn beperkt tot: readiness score, totale minuten diepe slaap van de nacht tussen de twee dagen, hartslagvariabiliteit en hartslag in rust. De gaten in de grafiek kunnen verklaard worden door het intensieve proces van datacleaning. Het kan bijvoorbeeld voorkomen dat een gebruiker zijn smartwatch tijdelijk niet heeft gedragen, of dat de data van de smartwatch onvolledig was voor die dagen waarop de grafiek onderbroken wordt.

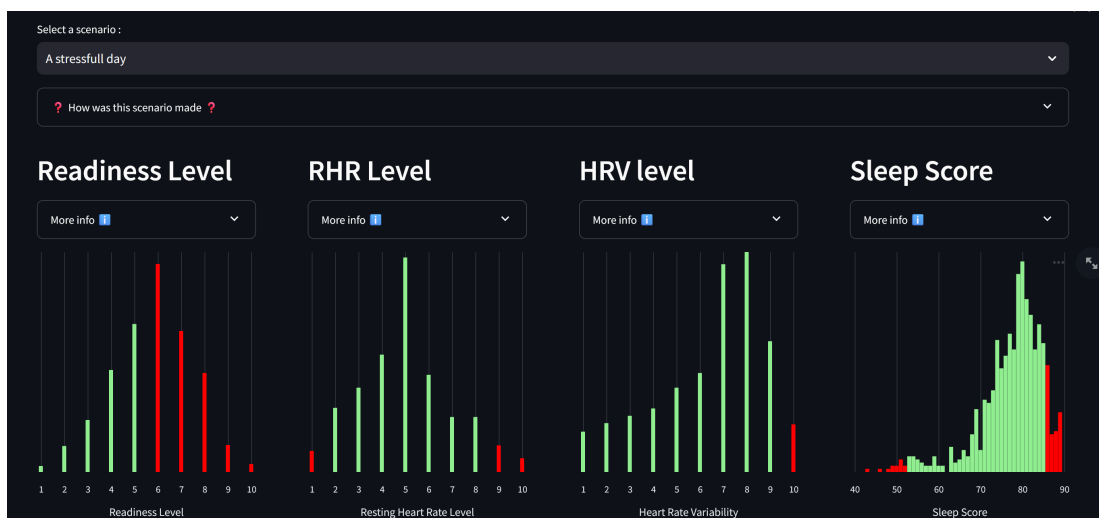


Figuur 6.3: Voorbeeld van een grafiek van twee complexere relaties

6.3 Introductie van scenario's

In een volgende sectie kan er dieper ingegaan worden op de dataset, met het oog op het gebruik van machine learning in de daaropvolgende sectie. Eerst worden een aantal scenario's gedefinieerd die in het dagelijks leven van iedereen kunnen voorkomen. Denk hierbij aan scenario's zoals: een stressvolle dag, een vakantiedag, een normale dag, een dag met slechte slaap en een dag waarop er intensief gesport werd. Hoe elk scenario is samengesteld, kan de gebruiker ontdekken in de visualisatie zelf (zie figuur 6.4).

Na het selecteren van een scenario kleuren de grafieken groen om het normale bereik van de parameters aan te geven. De rode kleur duidt aan dat de waarden buiten het normale bereik vallen voor de gegeven situatie, wat kan betekenen dat deze waarden zowel te laag als te hoog zijn. Naast de getoonde parameters op de figuur, zijn er nog vier andere parameters: minuten diepe slaap, totale slaapduur, totale tijd zittend doorgebracht, en *heart rate load* (trainingsbelasting over geheel de dag). Samen vormen deze acht parameters de datapunten van een bepaalde dag, waarmee een voorspelling kan worden gedaan voor het readiness score van de volgende dag. Indien de gebruiker een van de parameters niet begrijpt, is er een sectie met meer informatie beschikbaar waar hij kan lezen over de betreffende parameter.



Figuur 6.4: Verschillende scenario's en hun invloed op de parameters

6.4 Inputwaarden voor voorspelling

Wanneer de gebruiker voldoende heeft geleerd over elk scenario en de invloed van deze scenario's op de waarden van de belangrijke parameters, kan hij input leveren voor een voorspelling met behulp van machine learning. Hierbij worden dezelfde scenario's gebruikt als waarmee de gebruiker eerder heeft gewerkt. Het geven van input, kan gezien worden in figuur 6.5.

Het doel van deze scenario's is om zo realistisch mogelijke input te genereren. Zonder enige medische kennis zou het echter lastig zijn om geen onlogische combinaties van parameters te produceren, wat de nauwkeurigheid van de voorspelling negatief zou beïnvloeden. Daarom helpen de scenario's de gebruiker om realistische en consistente parameterwaarden te selecteren, waardoor de voorspellingen betrouwbaarder worden.

Na het kiezen van een scenario heeft de gebruiker nog de vrijheid om de parameters aan te passen, zodat de input beter overeenkomt met de data van zijn eigen smartwatch. Het platform ondersteunt de gebruiker verder door een waarschuwing te geven als hij een ongebruikelijke combinatie van parameters invoert. Dit helpt om ervoor te zorgen dat de input realistisch blijft en de voorspellingen zo nauwkeurig mogelijk zijn. Deze functie is ook zichtbaar onderaan in de figuur 6.5.

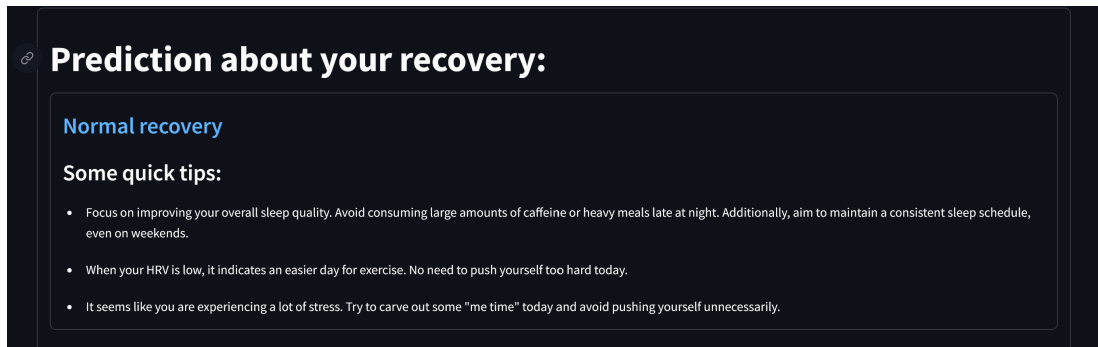
Figuur 6.5: Input-scherm voor machine learning voorspelling

6.5 Voorspelde herstelstatus en tips

Wanneer de gebruiker klaar is met het aanpassen van alle gewenste input voor een voorspelling, kan hij doorgaan naar het scherm dat hem een overzicht geeft van de gemaakte voorspelling door het machine learning-algoritme, zoals afgebeeld in figuur 6.6. Er zijn drie mogelijke voorspellingen met elk een eigen kleurcodering: ondermaats herstel (rood), normaal herstel (blauw) en bovengemiddeld herstel (groen). Deze labels komen overeen met de drie verschillende klassen van het readiness score in de dataset. Door middel van de kleur alleen kan in één oogopslag worden gezien wat de voorspelde status van het herstel is.

Daarnaast geeft de webapplicatie ook een paar tips mee om de herstelervaring van de gebruiker te verbeteren. Deze tips zijn gebaseerd op de parameters van de voorspelling waarop de gebruiker ondermaatse waarden heeft ingevuld. Deze tips worden weergegeven in het overzicht onder "herstel in meer detail" (zie figuur 6.7). Deze parameters met ondermaatse waarden zijn rood gekleurd om de gebruiker extra aandacht te geven.

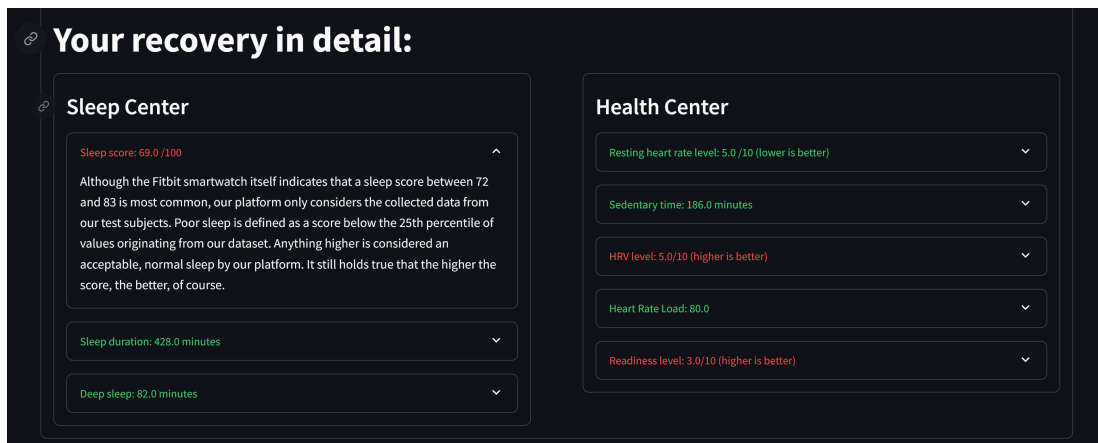
De laatste visualisatie (zie figuur 6.7) biedt een diepgaander inzicht in de verschillende componenten van het herstel. Voor elk component wordt een groene of rode kleur weergegeven. Groen



Figuur 6.6: Voorspelde status van het herstel met enkele tips om te verbeteren

geeft aan dat de gebruiker een bepaalde drempelwaarde voor de parameter heeft gehaald, terwijl rood aangeeft dat de gebruiker deze drempelwaarde niet heeft bereikt en hier nog verbetering mogelijk is.

Per component kan de gebruiker inzicht krijgen in hoe de drempelwaarde is berekend. Daarnaast is er extra uitleg beschikbaar over een component, of worden er tips gegeven om de score te verbeteren. Op deze manier krijgt een recreatieve sporter niet alleen meer inzicht in de belangrijkste componenten van het herstel, maar krijgt hij ook tips om actief zijn herstel te verbeteren. Zoals eerder benadrukt, is een adequaat herstel cruciaal om hoogwaardige training te blijven faciliteren.



Figuur 6.7: Meer informatie over de status van het herstel

Hoofdstuk 7

User Study

De gebruikersstudie valideert de gebruikerservaring van de ontworpen webapplicatie, zoals beschreven in Hoofdstuk 6. Het doel van deze studie is om te bepalen of de implementatie voldoet aan de vooraf gestelde eisen om een nauwkeurige voorspelling te doen over de trainingsstatus van een recreatieve atleet voor de volgende dag. Daarnaast moet de applicatie voldoende inzicht bieden in de samenstelling van de verschillende aspecten van het herstel, zodat de atleet de mogelijkheid heeft om zijn prestaties in de toekomst te verbeteren.

Hoewel de webapplicatie aan deze functionele vereisten voldoet, is het ook essentieel dat de gebruikerservaring (user experience) van de applicatie van hoge kwaliteit is. Een positieve gebruikerservaring is cruciaal, omdat het de waarschijnlijkheid vergroot dat recreatieve atleten de applicatie graag gebruiken en deze vaker en gedurende langere tijd blijven gebruiken.

Deze studie gebruikt een within-subject experimenteel ontwerp om een breed scala aan indrukken en gevoelens van gebruikers tijdens het gebruik van de applicatie te meten. Hierbij wordt onder andere gekeken naar de gebruiksvriendelijkheid, efficiëntie, en het algehele gebruikszelfvergenoegen. Door deze aspecten te evalueren, kan worden vastgesteld of de webapplicatie niet alleen functioneel maar ook aantrekkelijk en motiverend is voor de gebruikers. De resultaten van deze evaluatie zullen waardevolle inzichten bieden voor toekomstige verbeteringen en de verdere ontwikkeling van de applicatie.

7.1 Testprocedure

De volgende sectie beschrijft de procedure die elke deelnemer van de gebruiksstudie moet volgen bij het evalueren van de applicatie. Bij aankomst worden de deelnemers gevraagd een informed consent te ondertekenen. Op deze manier worden zij geïnformeerd over wat zij kunnen verwachten en geven zij hun toestemming voor het gebruik van hun data. Een belangrijk aspect dat in het informed consent wordt vermeld, is dat deelnemers te allen tijde mogen stoppen met de studie als zij dat willen.

Daarnaast ontvangen de deelnemers een kopie van het informed consent om mee naar huis te nemen. Dit stelt hen in staat hun rechten met betrekking tot de GDPR-wetgeving op elk moment uit te oefenen bij de verantwoordelijke instanties.

Na het ondertekenen van het informed consent, mogen de deelnemers plaatsnemen achter een laptop. Eerst wordt hen gevraagd een korte enquête in te vullen over hun leeftijd, sportachtergrond en kennis van smartwatches. Deze gegevens zijn waardevol voor het onderzoek, omdat ze kunnen helpen bij het verklaren van eventuele onverwachte resultaten.

Na de korte enquête kan de deelnemer beginnen met het testen van de applicatie. Om ervoor te

zorgen dat het testen voor elke deelnemer zo uniform mogelijk verloopt, heeft de onderzoeker een gestandaardiseerd scenario opgesteld. Dit scenario bevat een reeks stappen die de deelnemer in elke visualisatie moet doorlopen en vervolgens zelfstandig kan interpreteren. Een dergelijk scenario biedt een duidelijke leidraad voor wat de gebruiker in de applicatie kan doen en zorgt ervoor dat elke gebruiker de belangrijkste functionaliteiten test en ervaart.

De deelnemers hebben ook de vrijheid om meer te testen dan alleen het voorgeschreven scenario. Tijdens het testen kijkt de onderzoeker over de schouder van de deelnemer mee om visueel te controleren of de deelnemer het scenario volgt en om te observeren of er veel fouten worden gemaakt, wat kan wijzen op een slecht ontwerp. Hoewel dergelijke observaties niet worden genoteerd, zullen ze wel worden gebruikt om informele feedback te geven aan elke deelnemer nadat zij klaar zijn met het testen van de applicatie. Tijdens het testen wordt geen feedback gegeven om de resultaten niet te beïnvloeden.

De formele evaluatie van de applicatie zal echter uitsluitend gebaseerd zijn op de scores van de enquêtes die de deelnemers invullen na het testen. Deze enquêtes zijn bedoeld om kwantitatieve data te verzamelen over de gebruikerservaring en de tevredenheid met de applicatie.

Na het testen van de applicatie kreeg de testpersoon als eerste de SUS-vragenlijst, zie Hoofdstuk 7.2.2, om in te vullen op papier. Wanneer hij hiermee klaar was, ontving hij twee papieren van de UEQ-vragenlijst, zie Hoofdstuk 7.2.3. Het eerste blad was een kort instructieblad, dat ook aanwezig was in het sjabloon van de UEQ-vragenlijst. Het tweede blad bevatte de daadwerkelijke vragen van de UEQ-vragenlijst.

Tijdens het invullen van de vragenlijst zorgt de onderzoeker ervoor dat hij plaatsneemt tegenover de deelnemer, zodat deze niet het gevoel krijgt dat de onderzoeker meekijkt tijdens het invullen. Het voorkomen van deze indruk is belangrijk, omdat het meekijken ervoor zou kunnen zorgen dat deelnemers een druk ervaren om positievere resultaten in te vullen dan ze daadwerkelijk hebben ervaren.

Na het invullen van de laatste twee enquêtes werden deelnemers een laatste keer bedankt voor hun deelname aan de gebruikersstudie. Ze ontvingen hun kopie van de informed en mochten naar huis vertrekken.

7.2 Resultaten enquêtes

Aan deze gebruikersstudie nemen elf mensen vrijwillig deel. De participanten die eerder hadden deelgenomen aan de gegevensverzameling met de Fitbit-smartwatch, deden ook mee aan deze gebruikersstudie. De overige participanten werden gerekruteerd uit de kenniskring van Strava van de onderzoeker. Op deze manier was er enige zekerheid dat de deelnemers enige vorm van sport beoefenden en vertrouwd waren met de gegevens die een smartwatch kan verzamelen en weergeven.

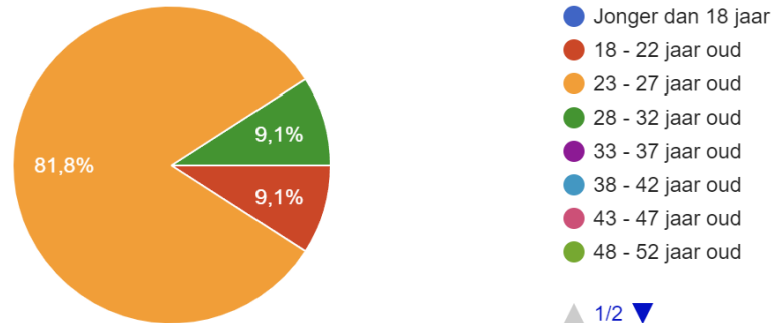
7.2.1 Algemene enquête

De volgende sectie bespreekt de resultaten van de eerste, algemene enquête die deelnemers moesten invullen voordat zij konden beginnen met het testen van de ontwikkelde webapplicatie.

Van de elf deelnemers waren er tien mannelijke en één vrouwelijke deelnemer. Alle deelnemers waren relatief jong, met geen enkele deelnemer ouder dan 35 jaar. De precieze verdeling van de leeftijden van de deelnemers is te zien in figuur 7.1.

Wat is je leeftijd?

11 antwoorden



Figuur 7.1: Verdeling leeftijden van de participanten

Hoewel de masterproef en de implementatie zich voornamelijk richten op wielrennen als duursport, werden andere duursporten en triatleten niet uitgesloten. Uit de enquête bleek dat zeven van de elf deelnemers voornamelijk recreatieve wielrenners waren. De overige vier personen beoefenden voornamelijk hardlopen als hun primaire duursport.

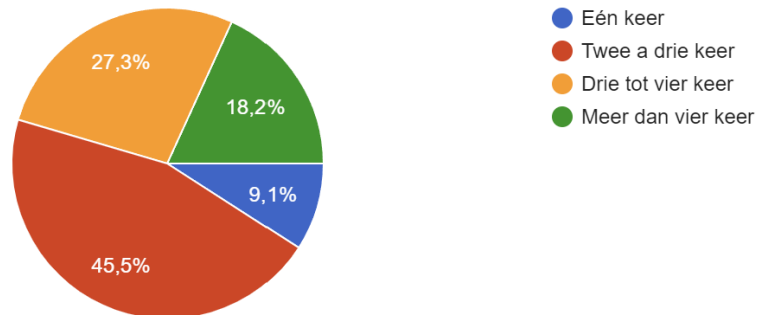
Tot slot werd ook geprobeerd om in te schatten hoe bekend de sporters zijn met de verschillende statistieken en parameters van een smartwatch, aangezien sommige van deze parameters zichtbaar zullen zijn in de webapplicatie. Dit werd gedaan aan de hand van twee vragen: 1) Hoe vaak gaat de persoon sporten per week. Dit kan een eerste indicatie geven hoe vaak de persoon gebruik kan maken van verschillende statistieken. 2) Hoelang heeft de persoon al een smartwatch. Dit kan aangeven of de gebruiker al ervaren is met het gebruik van een smartwatch of dat deze nog een beginner is. Hoewel beide opties niet met zekerheid kunnen vaststellen of de personen daadwerkelijk de meeste parameters van de smartwatch begrijpen, kunnen ze wellicht wel worden herkend. Personen die meer vertrouwd zijn met de parameters kunnen zich mogelijk ook meer vertrouwd en zelfverzekerd voelen wanneer er herkenningspunten zichtbaar zijn in een voor hen onbekende applicatie.

De antwoorden op beide vragen kunnen vervolgens worden gevonden in onderstaande figuren 7.2 en 7.3. Uit 7.2 valt op te merken dat het grootste gedeelte van alle deelnemers op een zeer regelmatige basis sport. Ongeveer 90 procent sport twee keer of vaker per week, sommige deelnemers sporten zelfs meer dan vier keer per week.

Uit figuur 7.3 blijkt dat het overgrote deel van alle deelnemers al bekend is met het gebruik van een smartwatch. Dit suggereert dat de meeste parameters in de visualisaties relatief vertrouwd zouden moeten aanvoelen voor de meeste gebruikers. Slechts een klein aantal deelnemers heeft recentelijk een smartwatch aangeschaft.

Hoe vaak beoefen je deze sport gemiddeld per week?

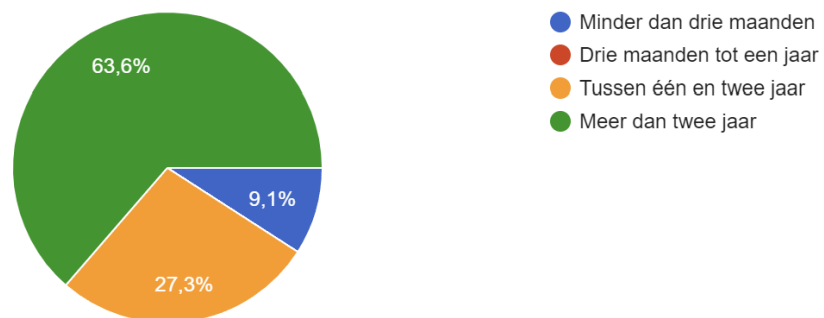
11 antwoorden



Figuur 7.2: Verdeling frequentie sporten van participanten

Hoelang maak je al gebruik van een smartwatch ?

11 antwoorden



Figuur 7.3: Verdeling ervaring smartwatchgebruik van participanten

7.2.2 System Usability Scale (SUS)

De System Usability Scale (SUS) is een vragenlijst die de gebruiksvriendelijkheid van een (web)applicatie kan meten aan de hand van tien verschillende vragen, waarbij vijf antwoordopties worden voorzien, variërend van "helemaal oneens" tot "helemaal eens". Door het beperkte aantal vragen, slechts tien, is deze methode zeer efficiënt om in korte tijd veel verschillende proefpersonen te testen. Dit is zowel voordelig voor onderzoekers als voor de proefpersonen, omdat ze niet belast worden met een onnodig lange enquête.

Een ander voordeel van de SUS-vragenlijst is dat deze resulteert in één enkele score op een schaal van 0 tot 100. Dit maakt het gemakkelijk om verschillende visualisaties of applicaties met elkaar te vergelijken op het gebied van gebruiksvriendelijkheid.

De SUS-score wordt aan de hand van de volgende stappen berekend:

- Schaalverdeling: Helemaal oneens = 1 punt, Helemaal eens = 5 punten
- Bereken de scores van alle oneven nummers door voor elke vraag $score_vraag - 1$ te doen en vervolgens alle resultaten op te tellen.

- Bereken dan de scores van alle even nummers door 5 - score_vraag te doen en vervolgens alle resultaten op te tellen.
- Tel de resultaten van de even en de oneven nummers samen en vermenigvuldig vervolgens met 2.5. Nu is een score op 100 verkregen.

De scores op een schaal van 0 tot 100 moeten licht anders worden geïnterpreteerd dan standaard beoordelingen op 100 volgens het framework van SUS. Een score van 51 of hoger mag dan wel nipt geslaagd zijn, maar de applicatie is dan nog niet erg gebruiksvriendelijk. Het wordt aangegeven dat verdere verbetering van de gebruiksvriendelijkheid een prioriteit moet zijn. Een score vanaf 68 duidt op een applicatie waarover de meningen gemiddeld zijn. Niemand stoort zich echt aan de applicatie, maar er zal ook niet erg lovend over worden gesproken. Pas vanaf een score van 80.3 wordt gesproken van een zeer goede applicatie die zeker zal worden aangeraden aan vrienden en kennissen.

Uit de verwerking van alle resultaten is geconcludeerd dat de applicatie, ontworpen voor deze masterproef, een SUS-score behaalde van 75.9. Daarnaast konden gebruikers ook nog een score op een schaal van 1 tot 10 geven als ze de website zouden aanbevelen aan andere personen. Daarop werd een gemiddelde score van 6.9 / 10 genoteerd. De SUS-score is een zeer goede, acceptabele score voor een eerste versie van de applicatie. Het lijkt erop dat de SUS-score ook goed overeenkomt met de gemiddelde aanbevelingsscore die deelnemers zelf hebben ingevuld.

7.2.3 User Experience Questionnaire (UEQ)

Naast SUS werd ook een User Experience Questionnaire (UEQ)¹ vragenlijst voorzien. Dit is ook een gestandaardiseerde vragenlijst die de gebruikerservaring van een (web)applicatie kan evalueren. Deze vragenlijst bevat echter meer vragen dan SUS en kan een breder scala aan gevoelens en indrukken meten. In totaal zijn er 26 vragen waarbij de invuller zijn eerste eigen gedachte invult, zonder veel te twijfelen. De UEQ-vragenlijst kan de volgende zes indrukken meten:

- Aantrekkelijkheid (*Attractiveness*): De algemene indruk van het product. Vinden gebruikers het leuk of net niet?
- Duidelijkheid (*Perspicuity*): Kan er gemakkelijk geleerd worden om het product te gebruiken?
- Efficiëntie (*Efficiency*): Kunnen gebruikers hun taak oplossen zonder onnodige moeite te doen? Reageert de applicatie snel?
- Afhankelijkheid (*Dependability*): Voelt de gebruikers zich in controle van de interacties? Is het veilig en voorspelbaar?
- Stimulatie (*Stimulation*): Is het opwindend en motiverend om de applicatie te gebruiken?
- Nieuwigheid (*Novelty*): Wordt een creatief design ontwikkeld? Wordt de interesse van de gebruikers vastgehouden?

Bij het berekenen van een score voor de UEQ-vragenlijst kan er geen algemene score worden berekend zoals bij de SUS-vragenlijst. Volgens de ontwerpers hield het geen steek om van zes verschillende aspecten slechts één score te berekenen. Daarom wordt voor elke categorie apart een score berekend. De ontwerpers van de vragenlijst bieden zelf ook een analyse tool aan waarbij alleen de antwoorden van de vragenlijsten moeten worden ingevuld. De tool berekent vervolgens zelf de scores, grafieken, enzovoort.

Belangrijk op te merken is de schaalverdeling van de scores. Voor elk aspect wordt een score tussen -3 en 3 berekend. Echter geven de auteurs aan dat het gebruik van deze grenzen een

¹<https://www.ueq-online.org/>

vertekend beeld kan geven. Uit onderzoek blijkt dat deelnemers aan vragenlijsten terughoudend zijn om de uitersten van de schaal te gebruiken. De auteurs stellen daarom voor om scores te laten zien op een schaal van -2 tot 2. Zo worden waarden tussen -0.8 en 0.8 beschouwd als een neutrale evaluatie. Wanneer de score hoger is dan 0.8, wordt dit gezien als een positieve evaluatie. Als deze 0.8-waarde echter op een maximale schaal van 3 zou worden geplaatst, zou op het eerste gezicht worden gedacht dat dit een vrij matige score zou zijn, wat in schril contrast staat met de positieve evaluatie.

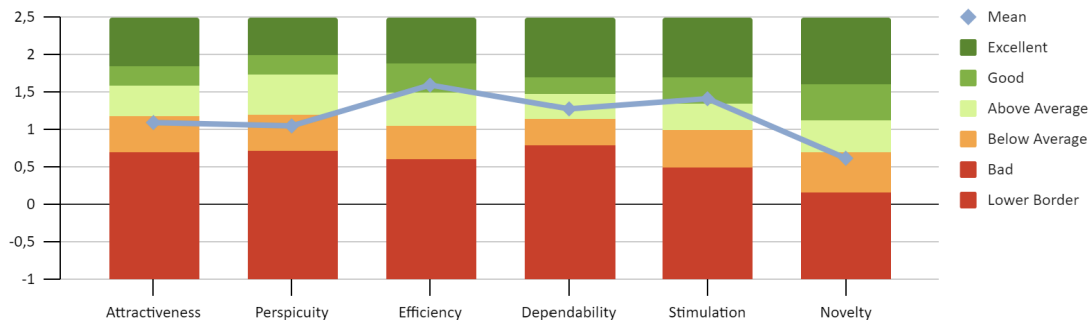
In onderstaande tabel zijn de scores voor elk aspect terug te vinden met betrekking tot de gemaakte implementatie van deze masterproef.

Aantrekkelijkheid	1.09
Duidelijkheid	1.05
Efficiëntie	1.59
Afhankelijkheid	1.27
Stimulation	1.41
Nieuwigheid	0.61

Tabel 7.1: UEQ-vragenlijst resultaten per aspect

Uit de bovenstaande resultaten valt op dat vijf van de zes aspecten een score van meer dan 0.8 hebben behaald, wat resulteert in een positieve waardering van alle respondenten. Met opmerkelijke uitschieters zijn de efficiëntie (1.59) en de stimulatie (1.41). Aan de andere kant heeft de nieuwigheid (0.61) een relatief bescheiden resultaat behaald.

In de UEQ-tool is ook een sectie te vinden met een benchmark. Deze functionaliteit stelt gebruikers in staat om de geteste applicatie te vergelijken met een uitgebreide dataset van applicaties die door UEQ worden beheerd.



Figuur 7.4: Resultaten tegenover de benchmark van UEQ

Uit de bovenstaande figuur blijkt dat de gemiddelde scores van alle applicaties niet uniform zijn. Bovendien lijken de voorgestelde scores van -0.8 en 0.8 voor een neutrale evaluatie optimistisch. Het merendeel van deze neutrale evaluaties valt voor de meeste aspecten in de lagere categorie. Bijgevolg kan een dergelijke neutrale evaluatie eerder als negatief worden beschouwd. Verder valt op dat de aspecten aantrekkelijkheid en duidelijkheid ongeveer een gemiddelde score hebben gekregen van alle respondenten, wat ook overeenkomt met het sentiment van de grenzen van UEQ. De aspecten efficiëntie, afhankelijkheid en stimulatie hebben zelfs een bovengemiddelde tot goede score behaald. Voor UEQ-normen betekent een goede score dat slechts tien procent van alle applicaties in de dataset een hogere score heeft. Dit is een mooi resultaat voor de aspecten efficiëntie en stimulatie.

Tot slot valt op dat de score voor nieuwigheid voor alle applicaties aanzienlijk lager ligt dan de gemiddelde waarden voor de andere aspecten. Het behaalde resultaat voor nieuwigheid (0.61), dat volgens de UEQ-standaard slechts matig is, blijkt een gemiddelde score te zijn voor alle applicaties. Dit suggereert dat de lagere score voor nieuwigheid (0.61) even hoog is als de score voor aantrekkelijkheid (1.09). Het zou dus zinvol zijn om in een volgende iteratie van de applicatie te focussen op het verbeteren van het aspect nieuwigheid. Ondanks de gemiddelde score heeft dit aspect nog aanzienlijk groeipotentieel.

Hoofdstuk 8

Future Work

In de volgende paragrafen zullen een aantal limitaties van de masterproef worden besproken. Vervolgens zal worden uitgelegd hoe deze limitaties in een volgende iteratie zouden verbeterd kunnen worden.

Ten eerste vertoont de dataset aanzienlijke beperkingen. Het aantal datapunten is beperkt door de korte duur van de datacollectie (3 maanden). Daarnaast was er slechts een klein aantal deelnemers, wat bijdroeg aan de beperkte omvang van de dataset. Door het kleine aantal deelnemers kan niet worden gesteld dat de dataset representatief is voor een gehele populatie gebruikers van smartwatches of duursporters. Verder valt op dat de verdeling van datapunten tussen deelnemers niet optimaal is; bijna de helft van de datapunten is afkomstig van slechts één persoon, wat de representativiteit van de dataset verder vermindert. Met meer tijd en budget zou een oplossing zijn om de datacollectie uit te breiden in zowel tijd als aantal deelnemers. Het zou bijvoorbeeld interessant zijn om personen gedurende een heel fietsseizoen te volgen. Dit zou ervoor zorgen dat er voor elke deelnemer voldoende datapunten aanwezig zijn in de dataset.

Naast de beperkingen van de dataset, valt op dat er een hoge nauwkeurigheid is (deels door de klasse-onbalans van de dataset), maar slechts beperkte resultaten voor de parameters sensitiviteit en precisie. Om het algoritme in een echte applicatie te kunnen gebruiken, zal het nodig zijn om deze parameters verder te verbeteren tot meer acceptabele niveaus. Dit probleem kan deels worden toegeschreven aan de kleine omvang van de dataset. Hoe meer kwalitatieve data wordt verzameld, hoe beter een machine learning-algoritme ervan kan leren. Daarnaast zijn niet alle verschillende machine learning-algoritmen getest. Het is mogelijk dat er een model bestaat dat beter presteert op de gegeven dataset.

Een andere beperking binnen deze masterproef is de manier waarop het herstel van een recreatieve atleet wordt ingeschat. De inschatting is uiteindelijk gebaseerd op een readiness score dat voortkomt uit de stress management score, ontworpen door Fitbit. Hierdoor ontstaat een vendor lock-in ten gunste van Fitbit. Ten eerste moet worden aangenomen dat Fitbit deze score op een degelijke en accurate manier berekent. Daarnaast mag de berekening van de score in toekomstige updates van het Fitbit-platform niet veranderen, aangezien dit een directe invloed zou hebben op eerder verzamelde data. Het is echter niet onmogelijk om de ontwikkelde implementatie te gebruiken voor andere platformen of merken van smartwatches, zolang de smartwatch dezelfde parameters kan meten en ook een readiness score kan bieden.

Hoofdstuk 9

Conclusies

9.1 Antwoord op research questions

In hoeverre kan een wearable een accurate schatting maken omtrent fysiologische stress van sportactiviteiten van een atleet? De fysiologische stress en de daaruit voortvloeiende trainingsbelasting kunnen in de eerste plaats worden berekend aan de hand van de hartslag tijdens de activiteit. Echter, vanwege de problemen met de smartwatches in deze masterproef, kan niet worden geconcludeerd dat wearables op een nauwkeurige manier de hartslag kunnen monitoren. Ten tweede kan de voorspelde readiness score van de volgende dag worden bekeken. Aangezien deze score meer omvat dan alleen de trainingsfactor, kan deze score slechts een inschatting geven van de totale fysiologische stress en niet specifiek de fysiologische stress door training kwantificeren.

Kan er met behulp van data van een wearable ingeschat worden wanneer een atleet voldoende hersteld is voor een nieuwe training? Uit de resultaten kan worden afgeleid dat er een relatief accurate voorspelling kan worden gemaakt omtrent de herstelstatus van een recreatieve atleet op basis van zijn smartwatchdata. Afhankelijk van het gekozen machine learning model varieert de nauwkeurigheid ruwweg tussen de 70 en 85 procent. De precisie is echter slechts ongeveer 40 procent, wat erop wijst dat er nog steeds een groot aantal *false positives* wordt gegenereerd. Dit duidt op een aanzienlijke ruimte voor verbetering.

Is het mogelijk om voor elke individuele atleet een gepersonaliseerd herstelvermogen te bepalen? Wegens de korte periode van datacollectie en het beperkte aantal datapunten per deelnemer, is het moeilijk om het herstelvermogen van een individu in deze dataset concreet in te schatten. Bovendien zijn er verschillende factoren die het herstel beïnvloeden, waaronder sommige die niet rechtstreeks gemeten kunnen worden door een smartwatch, zoals ziekte of alcoholgebruik. Het inschatten van het herstelvermogen vormt dan ook een beperking van zowel de dataset als de ontworpen webapplicatie.

Welke methodes zijn het meest gebruiksvriendelijk om data omtrent trainingen en herstel te visualiseren? Binnen deze masterproef werd een webapplicatie ontwikkeld voor duursporters om hun herstelstatus te voorspellen en de componenten ervan te visualiseren. Uiteindelijk werd de implementatie beperkt tot één ontwerp, waarop een *user study* werd uitgevoerd. Uit dit onderzoek kon niet worden afgeleid welke methoden superieur zijn aan andere. Wel bleek dat de huidige implementatie volgens de SUS- en UEQ-vragenlijsten een goede score behaalt. Maar er is nog ruimte voor verbetering.

9.2 Zelfreflectie

Bij het begin van het onderzoek had ik het gevoel dat er voldoende tijd was om deze masterproef te schrijven. Bijgevolg ben ik begonnen met een zo uitgebreid mogelijke literatuurstudie, die als een goede basis zou dienen voor zowel mijn onderzoek als mijn applicatie. Echter, door mijn persoonlijke interesse verloor ik mezelf enigszins in de interessante studies over wielrennen. Tegelijkertijd verschoof de focus van de masterproef beetje bij beetje. Aanvankelijk lag de focus op maximale prestaties in het wielrennen door herstel, supplementatie en training vanuit het perspectief van een amateur. Uiteindelijk zoomde de focus sterk in op het herstelaspect van de amateur. Dit betekende dat sommige literatuur achteraf gezien niet zo relevant was voor het onderzoek. Daarnaast had ik onvoldoende vooronderzoek gedaan naar het machine learning-aspect dat ik van plan was te implementeren in de applicatie. Hierdoor verloor ik enige tijd tijdens de implementatie van de applicatie en het opstellen van de resultaten. Tot slot vind ik dat ik kritischer had moeten zijn over mijn onderzoeksvragen. Pas na een aantal iteraties kwamen er duidelijkere onderzoeksvragen naar voren, wat pas gebeurde toen de literatuurstudie al vergevorderd was.

Over de manier waarop de data werd verzameld, ben ik tevreden. Door gebruik te maken van enkele smartwatches was het mogelijk de juiste doelgroep voor het onderzoek te selecteren. Er kan echter een argument worden gemaakt dat smartwatches niet altijd erg nauwkeurig zijn in vergelijking met de gouden standaard voor elk dataveld waarin wordt gemeten. Er bestaat echter geen goedkoper alternatief om op een dergelijke, onopvallende manier een recreatieve atleet de hele dag te monitoren. Achteraf werd ook geprobeerd om de verzamelde data uit te breiden met data die online beschikbaar is. Omdat het medische gegevens van deelnemers betreft, waren er echter nauwelijks datasets online beschikbaar die dezelfde parameters hadden gemeten. Daarnaast wil ik nog vermelden dat ik niet had verwacht zoveel problemen te ondervinden bij het verzamelen van de data van verschillende personen. Met vijf exact dezelfde modellen van een smartwatch had ik gedacht dat het vrij gemakkelijk zou zijn en dat elk horloge automatisch, zonder problemen, zijn metingen zou doen. Dat slechts de helft van de horloges correcte metingen heeft uitgevoerd, verbaast me nog steeds.

Wanneer ik mijn behaalde resultaten bekijk, ben ik persoonlijk toch enigszins tevreden. Hoewel mijn machine learning-prestaties niet baanbrekend zijn, heb ik niet het gevoel dat ik iets heb laten liggen. Ik kan mijn resultaten niet direct beïnvloeden, maar ik heb het maximale eruit gehaald met kruisvalidatie, hyperparameteroptimalisaties en het gebruik van SMOTE. Dit neemt niet weg dat de resultaten altijd beter kunnen. Voor de parameters sensitiviteit en precisie zijn er nog grote winsten te boeken. Daarnaast vind ik het persoonlijk een pluspunt dat er een implementatie uit de masterproef is voortgevloeid. Dit biedt de mogelijkheid voor iedereen om de applicatie uit te proberen en spelenderwijs iets te leren over de verschillende componenten van herstel, aan de hand van de voorspellingen en tips van de applicatie.

Uit het hele proces heb ik geleerd dat een goede planning belangrijk is, maar ook het vermogen om in te schatten hoelang een bepaalde taak zal duren. Tegelijkertijd is het belangrijk om op tijd te beseffen wanneer een bepaald onderdeel te veel tijd kost en nauwelijks nog van toegevoegde waarde zal zijn voor toekomstige onderdelen in de masterproef. Ook zal ik proberen in de toekomst nauwkeurig en duidelijke doelen te stellen waar ik concreet naartoe kan werken. Ik heb geleerd dat het me meer motivatie geeft om een doel te halen en af te werken, en tegelijkertijd zorgt het ervoor dat ik precies maak wat ik eerst in gedachten had. Zonder een concreet plan durf ik al eens op een zijspoor te geraken.

Bibliografie

- [1] S. Mitchell, J. Forsyth en M. S. Thompson, “Exploring amateur performance in athletic tests using wearable sensors,” in *2021 Systems and Information Engineering Design Symposium (SIEDS)*, IEEE, 2021, p. 1–6.
- [2] Business of Apps. “Strava Statistics 2024.” Accessed: 2024-06-08. (2024), adres: <https://www.businessofapps.com/data/strava-statistics/>.
- [3] K. Yoshida en K. Murao, “Load Position Estimation Method for Wearable Devices Based on Difference in Pulse Wave Arrival Time,” *Sensors*, jrg. 22, p. 1090, jan 2022. DOI: 10.3390/s22031090.
- [4] K. C. Vitale, R. Owens, S. R. Hopkins en A. Malhotra, “Sleep Hygiene for Optimizing Recovery in Athletes: Review and Recommendations,” DOI: 10.1055/a-0905-3103, 2019.
- [5] K. S. Seiler, G. Ø. Kjerland en S. Seiler, “Quantifying training intensity distribution in elite endurance athletes: is there evidence for an ”optimal” distribution?” *Scandinavian Journal of Medicine Science in Sports*, 2006. DOI: 10.1111/j.1600-0838.2004.00418.x.
- [6] G. Treff, K. Winkert, M. Sareban, M. Sareban, J. M. Steinacker en B. Sperlich, “The Polarization-Index: A Simple Calculation to Distinguish Polarized From Non-polarized Training Intensity Distributions.,” *Frontiers in Physiology*, 2019. DOI: 10.3389/fphys.2019.00707.
- [7] S. Sellés-Pérez, J. Fernández-Sáez en R. Cejuela, “Polarized and Pyramidal Training Intensity Distribution: Relationship with a Half-Ironman Distance Triathlon Competition,” *Journal of Sports Science and Medicine*, 2019. DOI: null.
- [8] T. Stöggl en B. Sperlich, “Polarized training has greater impact on key endurance variables than threshold, high intensity, or high volume training.,” *Frontiers in Physiology*, 2014. DOI: 10.3389/fphys.2014.00033.
- [9] T. Stöggl, T. Stöggl, T. Stöggl en G. Björklund, “High Intensity Interval Training Leads to Greater Improvements in Acute Heart Rate Recovery and Anaerobic Power as High Volume Low Intensity Training,” *Frontiers in Physiology*, 2017. DOI: 10.3389/fphys.2017.00562.
- [10] T. Stöggl en B. Sperlich, “The training intensity distribution among well-trained and elite endurance athletes,” *Frontiers in Physiology*, 2015. DOI: 10.3389/fphys.2015.00295.
- [11] A. Casado, F. González-Mohíno, J. M. González-Ravé en C. Foster, “Training Periodization, Methods, Intensity Distribution, and Volume in Highly Trained and Elite Distance Runners: A Systematic Review.,” *International Journal of Sports Physiology and Performance*, 2022. DOI: 10.1123/ijsp.2021-0435.
- [12] P. Bahenský en G. J. Grosicki, “Superior Adaptations in Adolescent Runners Using Heart Rate Variability (HRV)-Guided Training at Altitude.,” *Biosensors*, 2021. DOI: 10.3390/bios11030077.
- [13] G. Gremion, “The effect of stretching on sports performance and the risk of sports injury: A review of the literature,” *null*, 2005. DOI: null.
- [14] N. J. Tagliari en N. J. Tagliari, “Acute effect of stretching before power force exercises,” *International Physical Medicine Rehabilitation Journal*, 2019. DOI: 10.15406/ipmrj.2019.04.00206.

- [15] F. Esposito, E. Cè en E. Limonta, "Cycling efficiency and time to exhaustion are reduced after acute passive stretching administration," *Scandinavian Journal of Medicine Science in Sports*, 2012. DOI: 10.1111/j.1600-0838.2011.01327.x.
- [16] A. E. Wolfe, L. E. Brown, J. W. Coburn, R. D. Kersey en M. Bottaro, "Time Course of the Effects of Static Stretching on Cycling Economy," *Journal of Strength and Conditioning Research*, 2011. DOI: 10.1519/jsc.0b013e318234e55f.
- [17] L. M. Yamamoto, J. F. Klau, D. J. Casa, W. J. Kraemer, L. E. Armstrong en C. M. Maresh, "The effects of resistance training on road cycling performance among highly trained cyclists: a systematic review.," *Journal of Strength and Conditioning Research*, 2010. DOI: 10.1519/jsc.0b013e3181c86583.
- [18] K. Beattie, B. P. Carson, M. Lyons en I. C. Kenny, "The Effect of Maximal- and Explosive-Strength Training on Performance Indicators in Cyclists," *International Journal of Sports Physiology and Performance*, 2017. DOI: 10.1123/ijsp.2016-0015.
- [19] A. Sunde, Ø. Støren, M. Bjerkaas e.a., "Maximal Strength Training Improves Cycling Economy in Competitive Cyclists," *Journal of Strength and Conditioning Research*, 2010. DOI: 10.1519/jsc.0b013e3181aeb16a.
- [20] B. R. Rønnestad, E. A. Hansen en T. Raastad, "Strength training improves 5-min all-out performance following 185 min of cycling.," *Scandinavian Journal of Medicine Science in Sports*, 2011. DOI: 10.1111/j.1600-0838.2009.01035.x.
- [21] B. R. Rønnestad, E. A. Hansen en T. Raastad, "In-season strength maintenance training increases well-trained cyclists' performance.," *European Journal of Applied Physiology*, 2010. DOI: 10.1007/s00421-010-1622-4.
- [22] J. Robineau, N. Babault, J. Piscione, M. Lacombe en A.-X. Bigard, "Specific Training Effects of Concurrent Aerobic and Strength Exercises Depend on Recovery Duration.," *Journal of Strength and Conditioning Research*, 2016. DOI: 10.1519/jsc.0000000000000798.
- [23] P. Magill, T. Schwartz en M. Breyer, *Build Your Running Body*. 2017, p. 168–173.
- [24] W. D. McArdle, F. I. Katch en V. L. Katch, *Exercise Physiology: Nutrition, Energy, and Human Performance*, 7th. 2010, p. 138.
- [25] T. O. Bompa en C. A. Buzzichelli, *Periodization*. 2019, p. 20–22.
- [26] T. Purdom, L. Kravitz en K. e. a. Dokladny, "Understanding the factors that affect maximal fat oxidation," *Journal of the International Society of Sports Nutrition*, jrg. 15, nr. 1, p. 3, 2018. DOI: 10.1186/s12970-018-0207-1. adres: <https://doi.org/10.1186/s12970-018-0207-1>.
- [27] A. E. Jeukendrup, "Nutrition for endurance sports: Marathon, triathlon, and road cycling," *Journal of Sports Sciences*, 2011. DOI: 10.1080/02640414.2011.610348.
- [28] L. M. Burke en G. R. Cox, "Nutrition Strategies for Triathlon," *Triathlon Medicine*, 2020. DOI: 10.1007/978-3-030-22357-1_17.
- [29] A. E. Jeukendrup en A. E. Jeukendrup, "Training the Gut for Athletes," *Sports Medicine*, 2017. DOI: 10.1007/s40279-017-0690-6.
- [30] Gezond Leven. "Cafeïne." (N.D.), adres: <https://www.gezondleven.be/themas/voeding/focus-op-voeding-niet-voedingsstoffen/voedingsstoffen/caffe%C3%AFne>.
- [31] L. M. Burke, "Caffeine and sports performance.," *Applied Physiology, Nutrition, and Metabolism*, 2008. DOI: 10.1139/h08-130.
- [32] N. T. Jenkins, J. L. Trilk, A. Singhal, P. J. O'Connor, P. J. O'Connor en K. J. Cureton, "Ergogenic effects of low doses of caffeine on cycling performance.," *International Journal of Sport Nutrition and Exercise Metabolism*, 2008. DOI: 10.1123/ijsnem.18.3.328.
- [33] R. B. Kreider, D. S. Kalman, J. Antonio e.a., "International Society of Sports Nutrition position stand: safety and efficacy of creatine supplementation in exercise, sport, and medicine," *Journal of The International Society of Sports Nutrition*, 2017. DOI: 10.1186/s12970-017-0173-z.
- [34] D. L. Crisafulli, H. H. Buddhadev, L. R. Brilla e.a., "Creatine-electrolyte supplementation improves repeated sprint cycling performance: A double blind randomized control study," *Journal of The International Society of Sports Nutrition*, 2018. DOI: 10.1186/s12970-018-0226-y.

- [35] P. W. Mumford, W. C. Kephart, W. C. Kephart e.a., “Effect of 1-week betalain-rich beetroot concentrate supplementation on cycling performance and select physiological parameters.” *European Journal of Applied Physiology*, 2018. DOI: 10.1007/s00421-018-3973-1.
- [36] D. J. Muggeridge, C. C. Howe, O. Spendiff, C. R. Pedlar, P. E. James en C. Easton, “A single dose of beetroot juice enhances cycling performance in simulated altitude.” *Medicine and Science in Sports and Exercise*, 2014. DOI: 10.1249/mss.0b013e3182a1dc51.
- [37] K. E. Lansley, P. G. Winyard, S. J. Bailey e.a., “Acute Dietary Nitrate Supplementation Improves Cycling Time Trial Performance,” *Medicine and Science in Sports and Exercise*, 2011. DOI: 10.1249/mss.0b013e31821597b4.
- [38] M. V. Garnacho-Castaño, M. V. Garnacho-Castaño, G. Palau-Salvà e.a., “Effects of a single dose of beetroot juice on cycling time trial performance at ventilatory thresholds intensity in male triathletes.” *Journal of The International Society of Sports Nutrition*, 2018. DOI: 10.1186/s12970-018-0255-6.
- [39] D. P. Wilkerson, G. M. Hayward, S. J. Bailey, A. Vanhatalo, J. R. Blackwell en A. M. Jones, “Influence of acute dietary nitrate supplementation on 50 mile time trial performance in well-trained cyclists,” *European Journal of Applied Physiology*, 2012. DOI: 10.1007/s00421-012-2397-6.
- [40] W. H. Gurton, L. A. Gough, S. A. Sparks, M. A. Faghy en K. E. Reed, “Sodium Bicarbonate Ingestion Improves Time-to-Exhaustion Cycling Performance and Alters Estimated Energy System Contribution: A Dose-Response Investigation.” *Frontiers in Nutrition*, 2020. DOI: 10.3389/fnut.2020.00154.
- [41] B. Marrier, J. Robineau, J. Piscione e.a., “Supercompensation Kinetics of Physical Qualities During a Taper in Team Sport Athletes,” *International Journal of Sports Physiology and Performance*, 2017. DOI: 10.1123/ijsp.2016-0607.
- [42] Amstel. “Trainen met Fondo: Deel 1 - Waarom een trainingsschema?” (N.D.), adres: <https://www.amstel.nl/trainen-met-fondo-deel-1-waarom-een-trainingsschema>.
- [43] K. Tipton en R. Wolfe, “Exercise, protein metabolism, and muscle growth,” *Int J Sport Nutr Exerc Metab*, jrg. 11, nr. 1, p. 109–32, 2001. DOI: 10.1123/ijsnem.11.1.109.
- [44] TrainingPeaks. “The Endurance Athlete’s Guide to Protein.” (2022), adres: <https://www.trainingpeaks.com/blog/the-endurance-athlete-s-guide-to-protein/#:~:text=For%20most%20active%20individuals%20not,daily%20caloric%20requirements%20from%20protein..>
- [45] A. Mika, P. Mika, B. Fernhall en V. B. Unnithan, “Comparison of recovery strategies on muscle performance after fatiguing exercise.” *American Journal of Physical Medicine Rehabilitation*, 2007. DOI: 10.1097/phm.0b013e31805b7c79.
- [46] P. Weerapong, P. A. Hume en G. S. Kolt, “The Mechanisms of Massage and Effects on Performance, Muscle Recovery and Injury Prevention,” *Sports Medicine*, 2005. DOI: 10.2165/00007256-200535030-00004.
- [47] N. A. Martin, R. F. Zoeller, R. J. Robertson en S. M. Lephart, “The comparative effects of sports massage, active recovery, and rest in promoting blood lactate clearance after supramaximal leg exercise.” *Journal of Athletic Training*, 1998. DOI: null.
- [48] M. Hirshkowitz, K. Whiton, S. M. Albert e.a., “National Sleep Foundation’s sleep time duration recommendations: methodology and results summary,” *Sleep Health*, jrg. 1, nr. 1, p. 40–43, mrt 2015, Epub 2015 Jan 8. DOI: 10.1016/j.sleh.2014.12.010.
- [49] S. S. H. Roberts, W.-P. Teo, B. Aisbett en S. A. Warmington, “Extended Sleep Maintains Endurance Performance Better than Normal or Restricted Sleep,” *Medicine & Science in Sports & Exercise*, jrg. 51, nr. 12, p. 2516–2523, dec 2019. DOI: 10.1249/MSS.0000000000002071.
- [50] C. D. Mah, K. Mah, K. E. Mah, K. E. Mah, E. J. Kezirian en W. C. Dement, “The Effects of Sleep Extension on the Athletic Performance of Collegiate Basketball Players,” *Sleep*, 2011. DOI: 10.5665/sleep.1132.
- [51] M. D. Milewski, D. L. Skaggs, G. A. Bishop e.a., “Chronic lack of sleep is associated with increased sports injuries in adolescent athletes,” *Journal of Pediatric Orthopedics*, jrg. 34, nr. 2, p. 129–133, mrt 2014. DOI: 10.1097/BPO.0000000000000151.

- [52] S. Taheri, L. Lin, D. Austin, T. Young en E. Mignot, “Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index,” *PLoS Medicine*, jrg. 1, nr. 3, e62, dec 2004, Epub 2004 Dec 7. DOI: 10.1371/journal.pmed.0010062.
- [53] A. V. Nedeltcheva, J. M. Kilkus, J. Imperial, K. Kasza, D. A. Schoeller en P. D. Penev, “Sleep curtailment is accompanied by increased intake of calories from snacks,” *American Journal of Clinical Nutrition*, jrg. 89, nr. 1, p. 126–133, jan 2009, Epub 2008 Dec 3. DOI: 10.3945/ajcn.2008.26574.
- [54] P. S. Hogenkamp, E. Nilsson, V. C. Nilsson e.a., “Acute sleep deprivation increases portion size and affects food choice in young men,” *Psychoneuroendocrinology*, jrg. 38, nr. 9, p. 1668–1674, sep 2013, Epub 2013 Feb 18. DOI: 10.1016/j.psyneuen.2013.01.012.
- [55] S. M. Schmid, M. Hallschmid, K. Jauch-Chara e.a., “Disturbed glucoregulatory response to food intake after moderate sleep restriction,” *Sleep*, jrg. 34, nr. 3, p. 371–377, mrt 2011. DOI: 10.1093/sleep/34.3.371.
- [56] S. Cohen, W. J. Doyle, C. M. Alper, D. Janicki-Deverts en R. B. Turner, “Sleep habits and susceptibility to the common cold,” *Archives of Internal Medicine*, jrg. 169, nr. 1, p. 62–67, jan 2009. DOI: 10.1001/archinternmed.2008.505.
- [57] A. A. Prather, D. Janicki-Deverts, M. H. Hall en S. Cohen, “Behaviorally Assessed Sleep and Susceptibility to the Common Cold,” *Sleep*, jrg. 38, nr. 9, p. 1353–1359, sep 2015. DOI: 10.5665/sleep.4968.
- [58] S. H. Onen, A. Alloui, A. Gross, A. Eschallier en C. Dubray, “The effects of total sleep deprivation, selective sleep interruption and sleep recovery on pain tolerance thresholds in healthy subjects,” *Journal of Sleep Research*, jrg. 10, nr. 1, p. 35–42, mrt 2001. DOI: 10.1046/j.1365-2869.2001.00240.x.
- [59] T. Roehrs, V. Timms, A. Zwyghuizen-Doorenbos en T. Roth, “Sleep extension in sleepy and alert normals,” *Sleep*, jrg. 12, nr. 5, p. 449–457, okt 1989. DOI: 10.1093/sleep/12.5.449.
- [60] G. Borg, *Borg’s perceived exertion and pain scales*. Human kinetics, 1998.
- [61] R. H. Morton, J. R. Fitz-Clarke en E. W. Banister, “Modeling human performance in running,” *Journal of applied physiology*, jrg. 69, nr. 3, p. 1171–1177, 1990.
- [62] TrainingPeaks. “Training with TSS vs. HR-TSS: What’s the Difference?” (N.D.), adres: <https://www.trainingpeaks.com/learn/articles/training-with-tss-vs-hrtss-whats-the-difference/>.
- [63] Q. Bi, K. E. Goodman, J. Kaminsky en J. Lessler, “What is Machine Learning? A Primer for the Epidemiologist,” *American Journal of Epidemiology*, jrg. 188, nr. 12, p. 2222–2239, dec 2019. DOI: 10.1093/aje/kwz189.
- [64] W. Buntine en T. Niblett, “A Further Comparison of Splitting Rules for Decision-Tree Induction,” *Machine Learning*, jrg. 8, nr. 1, p. 75–85, 1992. DOI: 10.1023/A:1022686419106. adres: <https://doi.org/10.1023/A:1022686419106>.
- [65] H. Zhang en B. H. Singer, *Recursive partitioning and applications*. Springer Science & Business Media, 2010.
- [66] S. Ray, “A Quick Review of Machine Learning Algorithms,” *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, p. 35–39, 2019. adres: <https://api.semanticscholar.org/CorpusID:204229941>.
- [67] L. Breiman, “Random Forests,” *Machine Learning*, jrg. 45, nr. 1, p. 5–32, 2001. DOI: 10.1023/A:1010933404324. adres: <https://doi.org/10.1023/A:1010933404324>.
- [68] R. Hassanzadeh, M. Farhadian en H. Rafieemehr, “Hospital mortality prediction in traumatic injuries patients: comparing different SMOTE-based machine learning algorithms,” *BMC Medical Research Methodology*, jrg. 23, nr. 1, p. 101, 2023. DOI: 10.1186/s12874-023-01920-w. adres: <https://doi.org/10.1186/s12874-023-01920-w>.
- [69] M. Sewell, “Ensemble learning,” *RN*, jrg. 11, nr. 02, p. 1–34, 2008.
- [70] A. Natekin en A. Knoll, “Gradient boosting machines, a tutorial,” *Frontiers in neurobotics*, jrg. 7, p. 21, 2013.
- [71] C. Li, “A gentle introduction to gradient boosting,” jrg. 59, 2016. adres: http://www.ccs.neu.edu/home/vip/teach/MLcourse/4%5C_boosting/slides/gradient%5C_boosting.pdf.

- [72] Z. A. Ali, Z. H. Abduljabbar, H. A. Taher, A. B. Sallow en S. M. Almufti, "Exploring the power of eXtreme gradient boosting algorithm in machine learning: A review," *Academic Journal of Nawroz University*, jrg. 12, nr. 2, p. 320–334, 2023.
- [73] H. Feng, B. Yang, J. Wang e.a., "Identifying Malignant Breast Ultrasound Images Using ViT-Patch," *Applied Sciences*, jrg. 13, nr. 6, 2023, ISSN: 2076-3417. DOI: 10.3390/app13063489. adres: <https://www.mdpi.com/2076-3417/13/6/3489>.
- [74] Y. Zhuang, S. Chen, N. Jiang en H. Hu, "An Effective WSSENet-Based Similarity Retrieval Method of Large Lung CT Image Databases.," *KSI Transactions on Internet & Information Systems*, jrg. 16, nr. 7, 2022.
- [75] Y. Liu, J. Tian, R. Hu e.a., "Improved Feature Point Pair Purification Algorithm Based on SIFT During Endoscope Image Stitching," *Frontiers in Neurorobotics*, jrg. 16, p. 840594, feb 2022. DOI: 10.3389/fnbot.2022.840594.
- [76] S. Lu, B. Yang, Y. Xiao e.a., "Iterative reconstruction of low-dose CT based on differential sparse," *Biomedical Signal Processing and Control*, jrg. 79, p. 104204, 2023, ISSN: 1746-8094. DOI: <https://doi.org/10.1016/j.bspc.2022.104204>. adres: <https://www.sciencedirect.com/science/article/pii/S1746809422006589>.
- [77] N. V. Chawla, K. W. Bowyer, L. O. Hall en W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, jrg. 16, p. 321–357, 2002.
- [78] Q. Gu, Z. Cai, L. Zhu en B. Huang, "Data Mining on Imbalanced Data Sets," *2008 International Conference on Advanced Computer Theory and Engineering*, p. 1020–1024, 2008. adres: <https://api.semanticscholar.org/CorpusID:18186745>.
- [79] R. Ghorbani en R. Ghousi, "Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques," *IEEE Access*, jrg. 8, p. 67899–67911, 2020. adres: <https://api.semanticscholar.org/CorpusID:216042134>.
- [80] G. Kovács, "Smote-variants: A python implementation of 85 minority oversampling techniques," *Neurocomputing*, jrg. 366, p. 352–354, 2019, ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2019.06.100>. adres: <https://www.sciencedirect.com/science/article/pii/S0925231219311622>.
- [81] M. Umer, S. Sadiq, M. M. S. Missen e.a., "Scientific papers citation analysis using textual features and SMOTE resampling techniques," *Pattern Recognit. Lett.*, jrg. 150, p. 250–257, 2021. adres: <https://api.semanticscholar.org/CorpusID:237667244>.
- [82] D. Nanchen, "Resting heart rate: what is normal?" *Heart*, jrg. 104, nr. 13, p. 1048–1049, 2018.