

Data-driven onderzoek in nichesport: de schepcorner in het minivoetbal

Aantal woorden: 18.245

Simon Cosyns

Studentennummer: 02008275

Promotor: Dr. Bram Janssens

Masterproef voorgelegd voor het behalen van de graad Master of Science in Business Engineering - Data Analytics

Academiejaar 2024-2025

Vertrouwelijkheid en wetenschappelijke integriteit

De auteur en de promotor geven de toelating deze masterproef voor consultatie beschikbaar te stellen en delen van de masterproef te kopiëren voor persoonlijk gebruik. Elk ander gebruik valt onder de beperkingen van het auteursrecht, in het bijzonder met betrekking tot de verplichting de bron uitdrukkelijk te vermelden bij het aanhalen van resultaten uit de masterproef.

Ik verklaar dat het onderzoek verlopen is volgens de regels die gelden inzake wetenschappelijke en academische integriteit. Ik verklaar kennis te hebben genomen van, en gehandeld te hebben in overeenstemming met, de ethische code van de faculteit.

Gent, juni 2025

Simon Cosyns

Abstract

Sportanalytics is een snelgroeiende markt binnen het veld van data analytics (CAGR 20,6%¹), terwijl nichesporten wereldwijd aan populariteit winnen. Toch blijft wetenschappelijk onderzoek met data-analyse binnen nichesporten opvallend schaars. Deze masterproef wil deze leemte verkleinen door enerzijds de schepcorner in de nichesport minivoetbal diepgaand analytisch te bestuderen, en anderzijds een onderbouwde methodologie voor dataverzameling in nichesporten te ontwikkelen.

In het sporttechnische luik werd onderzocht welke impact de schepcorner heeft op wedstrijduitslagen en momentum, welke factoren bijdragen tot succes en wat het effect zou zijn van een aangepaste toekenningsregel. De analyses tonen aan dat de schepcorner een uitermate bepalende rol speelt: in meer dan 60% van de wedstrijden zou de uitslag wijzigen indien de schepcorners anders verliepen. Ook blijkt dat het scoren uit een schepcorner het momentum positief beïnvloedt. Daarnaast toont het onderzoek aan dat ervaring, en dus het werken met vaste duo's, de scoringskans verhoogt. Voor beginnende ploegen en teams op lager niveau blijkt het aangewezen dat de koper zich verder naar achter positioneert. Naarmate de technische precisie van de schepper toeneemt, verhoogt de scoringskans wanneer de koper zich dichterbij het doel en het strafschopgebied positioneert.

Wat betreft de methodologie van dataverzameling blijkt dat papieren enquêtes, ondanks hun praktische beperkingen, in dit onderzoek leidden tot meer accurate en vollediger ingevulde gegevens dan hun online tegenhangers. Dit werpt vragen op over de haalbaarheid en effectiviteit van online surveys binnen nichesportonderzoek. Daarnaast toont een vergelijking van verschillende enquête-opmaken aan dat de opbouw van de vragenlijst een rol speelt in de datakwaliteit. Een structuur waarbij eerst algemene en pas daarna repetitieve vragen worden gesteld, bevordert de kwaliteit van de antwoorden. Ook al kan de vragenlijst dan niet chronologisch ingevuld worden.

¹ <https://www.grandviewresearch.com/industry-analysis/sports-analytics-market>

Summary

This thesis investigates the tactical and statistical importance of the *schepcorner*, a unique set piece in *minivoetbal*, a Flemish variant of indoor five-a-side football. While the field of sports analytics has seen rapid growth in recent years, most academic attention remains focused on mainstream sports. In contrast, niche sports continue to be underrepresented in scholarly literature. This study addresses that gap by offering a data-driven analysis of *minivoetbal*, with a specific focus on the *schepcorner*, and by developing a replicable methodology for data collection in lesser-studied sporting contexts.

Minivoetbal distinguishes itself from other small-sided football formats through specific rules such as the prohibition on goalkeepers using their hands and the use of smaller goals (2.5 meters wide by 1 meter high). One of its most defining features is the *schepcorner*, a standardized play awarded after four accumulated corner points or as a penalty for certain fouls. During a *schepcorner*, the ball is placed near the goal, and one player (the lobber) scoops it to a teammate (the header), who attempts a header toward an undefended goal. No defenders are involved, and strict rules apply regarding positioning and execution.

Despite the strong regional popularity of *minivoetbal*, which boasts over 18.000 registered players and more than 13.000 official matches per year in Flanders, the sport has received virtually no attention in international academic databases. Even basic search terms such as ‘minivoetbal’ or ‘schepcorner’ yield no relevant scientific studies. Furthermore, the English term ‘minifootball’ often refers to entirely different formats, further complicating literature searches. This academic neglect is problematic because niche sports are steadily growing in participation and societal impact. Voetbal Vlaanderen, the official federation overseeing the sport, is actively seeking ways to lower barriers to entry, particularly by clarifying complex rules like the *schepcorner*. This thesis responds to those needs by aiming to demystify the *schepcorner* and explore its strategic role in the game.

The research has two main objectives. First, it aims to assess the *schepcorner's* impact on match outcomes and momentum, evaluate the effects of a hypothetical rule change, and identify key tactical factors associated with success. Second, it seeks to develop and evaluate a methodological framework for collecting data in niche sports, where no central databases or automated tracking systems are available.

To achieve these goals, the study combines three methods of data collection: paper-based surveys completed in real time by substitute players, a less successful online survey, and video analyses of matches at the national level. In total, 603 *schepcorners* from 231 matches were documented. The data were cleaned and standardized before being analyzed through both descriptive statistics and machine learning models (logistic regression, Random Forest, and XGBoost). Additional analyses focused on the importance of the *schepcorner*, momentum shifts, hypothetical rule changes, and variable significance.

The findings demonstrate that the *schepcorner* is not merely a formal aspect of the game, but a decisive one. In over 25% of matches, altering the result of just one *schepcorner* would change the final outcome. If all *schepcorners* were hypothetically reversed, the outcome of over 60% of matches would differ. Scoring a *schepcorner* also correlates with increased momentum, whereas missing one does not systematically lead to a decline in performance.

Rule simulations reveal that changing the threshold for awarding a *schepcorner* from four to three corners would increase the total number of *schepcorners* by more than 50%, while raising the

threshold to five would result in a 25% decrease. Such adjustments could alter league rankings and affect competitive fairness, suggesting that even small rule modifications deserve careful empirical evaluation.

Regarding success factors, the quality of the lob is the most decisive element. When the header must move backwards or horizontally, the chance of scoring drops dramatically. In contrast, *schepcorners* where the header remains largely stationary are converted 88,6% of the time. Experience is another strong predictor. Teams that consistently use fixed lobber-header pairs, and players with higher individual or shared experience, achieve significantly better outcomes. Additionally, playing at a higher competition level correlates with both better lob execution and higher scoring efficiency. Interestingly, the preferred position of the header also shifts by level: lower-level teams tend to position the header further back to allow more error tolerance, while higher-level teams place the header closer to the goal, requiring greater precision but yielding higher rewards.

This study also carries important methodological implications. While online surveys are often promoted as efficient tools, this research found them largely ineffective in the context of recreational indoor sports. Paper-based surveys achieved higher completion rates and more consistent data quality. The order of questions and the physical design of the survey form further impacted data completeness, suggesting that small structural choices can make a substantial difference.

The practical implications are twofold. Coaches and teams can use these insights to better train and position players during *schepcorners*, especially by working with consistent pairs and adjusting player positions based on skill level. Policymakers within Voetbal Vlaanderen can draw on the findings to evaluate potential rule changes and to better communicate the role of *schepcorners* to new players.

Nonetheless, the study has limitations. The dataset was based on a convenience sample, which overrepresents certain lower-level teams. The online and paper surveys were distributed at different times in the season, complicating direct comparisons. Predictive models without the quality of the lob data performed substantially worse, indicating a need for more granular pre-execution variables. Finally, this thesis did not include qualitative research into player or coach perceptions, which could enrich future studies.

In summary, this thesis offers the first data-driven investigation into the *schepcorner* in *minivoetbal*. It shows that this unique set piece plays a critical role in determining match outcomes, that its execution can be optimized through tactical adjustments and experience, and that even niche sports benefit from rigorous analytical frameworks. At the same time, it highlights the challenges and opportunities in conducting empirical research in underexplored sports domains, paving the way for future work in this expanding field.

Woord vooraf

Met het schrijven van deze masterproef rond ik mijn opleiding tot handelsingenieur met afstudeerrichting Data Analytics aan de Universiteit Gent af. Dit onderzoek markeert het einde van een intensieve, leerrijke en uitdagende studieloopbaan. Deze masterproef had echter nooit gerealiseerd kunnen worden zonder de steun en hulp van enkele mensen die ik hierbij graag wil bedanken.

In de eerste plaats wil ik Dr. Bram Janssens oprecht bedanken. Zijn begeleiding, constructieve feedback en inhoudelijke suggesties hielpen mij om een helder en haalbaar onderwerp af te bakenen en zorgden ervoor dat ik steeds richting vond tijdens het schrijfproces.

Daarnaast ben ik bijzonder dankbaar aan Willy Van Leirberghe en Kevin Van De Wiele, de drijvende krachten achter het minivoetbal. Willy speelde een cruciale rol bij het faciliteren van de dataverzameling, onder meer door zaalwachters te overtuigen om enquêtes uit te delen en door ploegen actief aan te sporen om deel te nemen via herhaalde communicatie. Kevin bood waardevol inzicht in hoe Voetbal Vlaanderen naar het minivoetbal kijkt, ondersteunde het onderzoek vanuit zijn rol binnen de organisatie en vulde zelf ook enquêtes in. Zonder hun enthousiaste betrokkenheid en praktische hulp zou dit onderzoek onmogelijk zijn geweest.

Ook wil ik alle ploegen bedanken die hun tijd namen om de enquête in te vullen en zo bijdroegen aan het mogelijk maken van deze studie.

Verder wil ik mijn familie en vrienden bedanken voor hun voortdurende steun, luisterend oor en motiverende woorden tijdens de soms drukke werkperiodes. Hun aanwezigheid maakte een groot verschil.

Tot slot een knipoog naar mijn teamgenoten, die het steeds met de glimlach toelieten dat ik het missen van een schepcorner toeschreef aan wetenschappelijk onderzoek. Bedankt voor het leuke seizoen.

Simon Cosyns
Gent, juni 2025

Inhoudsopgave

Vertrouwelijkheid en wetenschappelijke integriteit.....	II
Abstract	III
Summary	IV
Woord vooraf	VI
Lijst van afkortingen	X
Lijst van tabellen.....	X
Lijst van figuren	XI
1 Introductie	1
2 Literatuurstudie.....	3
2.1 De schepcorner: inzichten uit andere sporten.....	3
2.1.1 Druk als bepalende factor	4
2.1.2 De rol van ervaring en vermoeidheid	4
2.1.3 Technische vaardigheid als verklarende factor	4
2.1.4 Methodologische benaderingen	4
2.1.5 Samenvattende tabel	5
2.2 Onderzoek in nichesporten	6
2.2.1 Vergelijking tussen papieren en online enquêtes	7
2.2.2 Survey Design	8
2.3 Conclusie literatuurstudie	10
3 Methodologie	12
3.1 Dataverzameling.....	12
3.1.1 Ontwikkeling van de vragenlijst	12
3.1.2 Online enquête.....	12
3.1.3 Papieren enquête	12
3.1.4 Video-observatie van wedstrijden	13
3.1.5 Controle van datakwaliteit	13
3.1.6 Parameters uit enquête.....	14
3.2 Data preprocessing.....	15
3.2.1 Data cleaning	15
3.2.2 Transformaties	15
3.2.3 Feature engineering	16
3.3 Model selection	17
3.4 Model training en evaluatie	18
3.5 Exploratieve analyses	19

3.5.1 Impactanalyse.....	19
3.5.2 Momentumanalyse	20
3.5.3 Effect van een hypothetische regelwijziging	20
3.5.4 Analyse van variabelen	21
4 Resultaten.....	22
4.1 Resultaten enquête en dataverzameling	22
4.1.1 Online enquête	22
4.1.2 Papieren enquête	23
4.1.3 Video analyse.....	23
4.1.4 Data controle	24
4.2 Impact.....	25
4.3 Momentum.....	26
4.4 Impact regelwijziging.....	28
4.5 Belangrijke factoren achter het succes van de schepcorner.....	31
4.5.1 Vermoeidheid	32
4.5.2 Positionele kenmerken	33
4.5.3 Schepkwaliteit	36
4.5.4 Ervaring.....	38
4.6 Voorspellende modellen voor schepcorner succes.....	39
4.6.1 Voorspellende modellen met schepkwaliteit-indicatoren	39
4.6.2 Voorspellende modellen zonder schepkwaliteit-indicatoren	44
5 Discussie	46
5.1 Interpretatie van de resultaten	47
5.1.1 Dataverzameling en enquêtes.....	47
5.1.2 Impact.....	47
5.1.3 Momentum.....	48
5.1.4 Bepalende factoren voor schepcornersucces	48
5.2 Praktische implicaties van de onderzoeksresultaten	49
5.2.1 Dataverzameling in de nichesport.....	49
5.2.2 De schepcorner in het minivoetbal	50
6 Beperkingen en toekomstig onderzoek	51
Literatuurlijst	XII
Appendix.....	XVI
A.1 Papieren enquête versie 1	XVI
A.2 Papieren enquête versie 2	XX
A.3 Hyperparameters Random Forest.....	XXIV

A.4 Hyperparameters XGBoost	XXV
A.5 Posities per niveau	XXVI
A.6 Visualisaties importance XGBoost	XXVII
A.7 Confusion Matrix.....	XXVIII
A.8 Visualisaties importance zonder schepkwaliteit.....	XXIX

Lijst van afkortingen

AUC	Area Under the Receiver Operating Characteristic Curve
ANOVA	ANalysis Of VAriance
LR	Logistic Regression
RF	Random Forest
ROC	Receiver Operating characteristic Curve
XGB	eXtreme Gradient Boosting
XGBoost	eXtreme Gradient Boosting

Lijst van tabellen

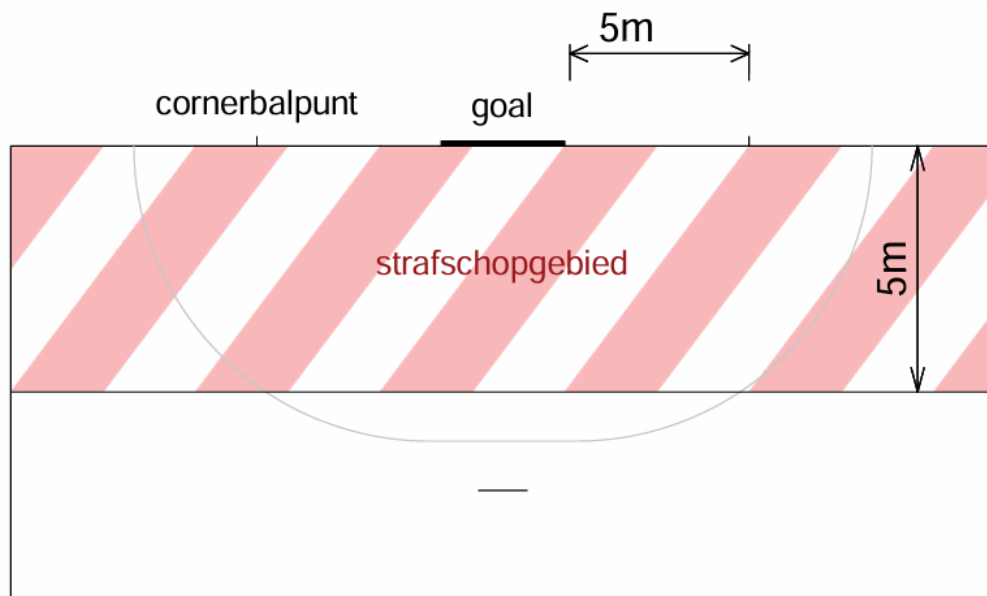
Tabel 1: Google Scholar hits	3
Tabel 2: Samenvattende tabel	5
Tabel 3: Wedstrijdparameters.....	14
Tabel 4: Schepcornerparameters	14
Tabel 5: Geconstrueerde variabelen	16
Tabel 6: Overzicht dataverzamelmethode.....	23
Tabel 7: Overeenstemming controle en zelfgerapporteerde enquêtes	24
Tabel 8: Impact van schepcorners.....	25
Tabel 9: Vergelijking momentum voor en na schepcorner	27
Tabel 10: Impact van regelwijziging	28
Tabel 11: Puntenverschil en eindklassering bij regelwijziging	30
Tabel 12: Gegeven per competitieniveau	31
Tabel 13: Schepcornersucces verdeeld op bankbezetting	33
Tabel 14: Scorepercentage verdeeld op schepkwaliteit-indicatoren.....	36
Tabel 15: Scorepercentage verdeeld per ervaringsniveau.....	38
Tabel 16: Evaluatie van voorspellende modellen.....	39
Tabel 17: Regressie-output	41
Tabel 18: Evaluatie van voorspellende modellen zonder schepkwaliteit-indicatoren	44
Tabel 19: Regressie-output zonder schepkwaliteit-indicatoren	45
Tabel 20: Hyperparameters Random Forest	XXIV
Tabel 21: Geselecteerde hyperparameters voor Random Forest modellen.....	XXIV
Tabel 22: Hyperparameters XGBoost.....	XXV
Tabel 23: Geselecteerde hyperparameters voor XGBoost modellen.....	XXV
Tabel 24: Voorspellingsresultaten per model	XXVIII

Lijst van figuren

Figuur 1: Speelveld minivoetbal met afmetingen	1
Figuur 2: Algemeen momentum na schepcorners	26
Figuur 3: Vergelijking momentum na gemiste vs. gescoorde schepcorners.....	27
Figuur 4: Extra of minder doelpunten op een seizoen per team bij regelwijziging	29
Figuur 5: Aantal schepcorners per kwart met scorepercentages	32
Figuur 6: Schets van mogelijke posities van de koper.....	33
Figuur 7: Aantal schepcorners per positie van de koper met de scorepercentages.....	34
Figuur 8: Scorepercentage en verdeling van posities over de verschillende competitieniveaus	35
Figuur 9: Schepkwaliteit vergeleken op competitieniveau	37
Figuur 10: ROC curves van verschillende voorspellende modellen	40
Figuur 11: Feature importance van het Random Forest model	42
Figuur 12: Cross-Validated permutation feature importance.....	43
Figuur 13: Posities per niveau	XXVI
Figuur 14: XGBoost feature importance	XXVII
Figuur 15: XGBoost permutation importance	XXVII
Figuur 16: RF zonder schepkwaliteit feature importance	XXIX
Figuur 17: RF zonder schepkwaliteit permutation importance	XXIX
Figuur 18: XGB zonder schepkwaliteit feature importance	XXX
Figuur 19: XGB zonder schepkwaliteit permutation importance.....	XXX

1 Introductie

In het veld van sportanalytics vormen gestandaardiseerde spelmomenten al langer een waardevolle bron van onderzoek. Toch blijven bepaalde sporten, zoals minivoetbal, ondervertegenwoordigd in de wetenschappelijke literatuur. Deze masterproef focust op een uniek, gestandaardiseerd moment binnen het minivoetbal: de cornerbal.



Figuur 1: Speelveld minivoetbal met afmetingen

Minivoetbal is een variant op voetbal waarbij indoor vijf tegen vijf wordt gespeeld. De sport onderscheidt zich van andere zaalvoetbalvormen door unieke regels, zoals het verbod voor de keeper om de bal met de handen te spelen en de kleinere doelen (2,5 m breed x 1 m hoog). Een andere specifieke spelregel binnen het minivoetbal is de cornerbal, beter bekend onder de benaming schepcorner. Aangezien traditionele hoekschoppen minder effectief zijn in dit format, kent men per hoekschoop één cornerpunt toe. Na vier verzamelde cornerpunten krijgt het team een cornerbal en worden de cornerpunten opnieuw op nul gezet. Daarnaast kan een cornerbal ook direct worden toegekend bij een overtreding waarbij een reële doelkans buiten het strafschopgebied wordt ontnomen. De uitvoering van een cornerbal verloopt als volgt: de bal wordt links of rechts van het doel op een cornerbalpunt, vijf meter van de dichtstbijzijnde paal geplaatst, waarna een eerste speler, de schepper, de bal met één beweging richting een tweede speler, de koper, speelt. Deze kopt de bal vervolgens in één keer richting het onverdedigde doel. Een cornerbal is een doelpunt als de bal gescoord wordt en de koper het strafschopgebied niet betreedt. Verder is er bij de cornerbal geen tegenstander betrokken. Zij moeten zich verwijderen van de helft waarop de cornerbal plaatsvindt. Voor verdere details over de regels van het minivoetbal wordt verwezen naar het officiële [bondsreglement](https://belgianfootball.s3.eu-central-1.amazonaws.com/s3fs-public/rbfa/docs/pdf/reglement/bondsreglement_reglement_federal/KBVB_bondsreglement_Boek_M.pdf)². Een visuele demonstratie van de cornerbal is te raadplegen via deze [video](https://youtu.be/RDmAPL78dn0?si=mfOYzfXWM72kC-d&t=342)³

² https://belgianfootball.s3.eu-central-1.amazonaws.com/s3fs-public/rbfa/docs/pdf/reglement/bondsreglement_reglement_federal/KBVB_bondsreglement_Boek_M.pdf

³ <https://youtu.be/RDmAPL78dn0?si=mfOYzfXWM72kC-d&t=342>

Minivoetbal wordt in Vlaanderen op ruime schaal beoefend, met 18.378 aangesloten leden en jaarlijks 13.176 officiële wedstrijden. Daarnaast bestaan er nationale en provinciale competities, elk met vier niveaus, en meer recreatieve regionale reeksen, ook wel kerncompetities genoemd. Deze kerncompetities zijn verspreid over heel Vlaanderen. Dit onderzoek richt zich voornamelijk op de Gentse kerncompetitie, die op haar beurt is onderverdeeld in vier competitieniveaus, en het hoogste nationale niveau.

Ondanks zijn stevige verankering in Vlaanderen is minivoetbal internationaal nauwelijks bekend. Hoewel de Engelse term ‘minifootball’ op het eerste gezicht een veelbelovende ingang voor academisch onderzoek lijkt, met ruim 2.700 resultaten op Google Scholar, blijkt bij nadere analyse dat de term vaak wordt gebruikt als synoniem voor zaalvoetbal, of voor een andere variant van voetbal die populair is in Oost-Europese landen en eveneens de naam ‘minifootball’ draagt. Dit zorgt voor verwarring, aangezien zaalvoetbal en het Oost-Europese minifootball verschillen van het minivoetbal. Zo worden die sporten onder meer gespeeld met een vaste keeper en grotere doelen, en ontbreken specifieke spelregels zoals de schepcorner. Daardoor zijn deze bronnen inhoudelijk niet representatief voor het minivoetbal zoals dat in Vlaanderen wordt gespeeld.

Een zoekopdracht op de Nederlandse term ‘minivoetbal’ toont onmiddellijk de beperkte wetenschappelijke aandacht voor deze sport: een zoekopdracht op deze term levert slechts een vijftigtal resultaten op. Bovendien blijken deze verwijzingen zelden over de sport zelf te gaan, maar eerder louter vermeldingen in het voorwoord van een werk, bijvoorbeeld als dankbetuiging aan een team. Dit gebrek aan inhoudelijke literatuur toont aan dat minivoetbal nauwelijks onderwerp is van academisch onderzoek. Ook de term ‘schepcorner’, een belangrijk en uniek aspect van deze sport, levert geen enkele academische bron op. Deze vaststellingen wijzen op een duidelijke lacune in het bestaande wetenschappelijke onderzoek naar zowel minivoetbal als de schepcorner in het bijzonder.

Deze masterproef speelt in op die leemte en beoogt een diepgaand inzicht te bieden in het fenomeen van de schepcorner. Deze studie onderzoekt (1) welke impact deze spelhervatting heeft op de wedstrijduitslag; (2) of het al dan niet scoren ervan momentum creëert; (3) wat het effect is van een regelwijziging; en (4) hoe een schepcorner optimaal kan worden uitgevoerd en welke factoren daarbij een rol spelen. Na overleg met Voetbal Vlaanderen blijkt dat deze vragen ook voor beleidsmakers en spelers bijzonder relevant zijn. De schepcorner is immers een minivoetbalspecifiek element dat, mede door zijn complexiteit, een drempel kan vormen voor nieuwe spelers. Dit onderzoek beoogt die drempel te verlagen en de sport toegankelijker te maken. Daarnaast wilt deze studie beleidsmakers een volledig beeld geven van de betekenis van deze stilstaande fase voor het minivoetbal en van de mogelijke gevolgen van regelwijzigingen.

Naast de praktische relevantie biedt dit onderzoek ook een methodologische meerwaarde. Het belang van deze studie reikt verder dan het minivoetbal alleen. Door een methodologisch kader te ontwikkelen voor dataverzameling binnen nichesporten, laat deze masterproef zien hoe minder populaire sporten datagedreven onderzocht kunnen worden.

De rest van deze masterproef is als volgt opgebouwd: [Hoofdstuk 2](#) biedt een overzicht van de literatuur over gelijkaardige gestandaardiseerde spelmomenten in populaire sporten en de dataverzameling in nichesporten. [Hoofdstuk 3](#) beschrijft de methodologie en het analysekader. [Hoofdstuk 4](#) presenteert de resultaten, gevolgd door een interpretatie in [Hoofdstuk 5](#). [Hoofdstuk 6](#) besluit met de beperkingen van het onderzoek en suggesties voor toekomstig werk.

2 Literatuurstudie

Een eenvoudige zoekopdracht op Google Scholar toont aan dat er, voor zover de onderzoeker kon vaststellen, geen academisch onderzoek bestaat naar minivoetbal vanuit een data-analytisch perspectief. Terwijl brede termen zoals 'data analytics' of 'sports analytics' honderdduizenden tot miljoenen treffers opleveren, blijft het aantal publicaties over 'minifootball analytics' uiterst beperkt (slechts 38 resultaten), die bovendien, zoals in de introductie aangehaald, betrekking hebben op andere sporten.

Term	Hits
Data analytics	3,370,000
Sports analytics	506,000
Soccer analytics	34,900
Minifootball analytics	38
Minivoetbal analytics	0

Tabel 1: Google Scholar hits

Deze asymmetrie weerspiegelt een bredere trend binnen sportonderzoek, waarin academische aandacht zich grotendeels beperkt tot mainstream sporten. Deze dominante sporten beschikken over grotere spelerspopulaties, sterkere institutionele omkadering en vooral: toegankelijke, gestandaardiseerde databanken. In tegenstelling hiermee ontbreekt in nichesporten zoals minivoetbal vaak de digitale infrastructuur voor systematische dataregistratie, wat empirisch onderzoek sterk bemoeilijkt. Daardoor blijven deze sporten grotendeels onderbelicht binnen sportanalytisch onderzoek, ondanks hun potentieel om waardevolle inzichten te bieden.

Dit kennishiaat vormt de aanleiding voor deze masterproef, die een dubbele focus hanteert. Enerzijds worden studies besproken over vergelijkbare, gestandaardiseerde spelmomenten uit andere sporten. Anderzijds wordt aandacht besteed aan methodologische inzichten uit nichesportonderzoek, met nadruk op dataverzameling via zelf ontwikkelde surveys. Deze brede benadering is noodzakelijk om zowel inhoudelijke als methodologische lacunes in de bestaande literatuur aan te pakken.

2.1 De schepcorner: inzichten uit andere sporten

Hoewel de schepcorner een unieke spelhervatting is binnen het minivoetbal, kan men deze tot op zekere hoogte gelijkstellen aan spelhervattingen uit andere sporten. Zo zijn er parallellen met penalty's in het veldvoetbal (Dohmen, 2008; Jordet et al., 2007) en field goals in American football (Clark et al., 2013). Net zoals bij de schepcorner gaat het in beide gevallen om een individueel moment binnen een teamsport, waarbij een speler onder hoge druk een gestandaardiseerde uitvoering moet leveren die direct impact kan hebben op het wedstrijdverloop. Ondanks het feit dat er bij penalty's een doelman aanwezig is en bij field goals een verdedigingsteam, blijft de essentie vergelijkbaar: de speler beschikt over een moment van relatieve isolatie om een technisch specifieke actie te voltooien, waarbij stress en ervaring vaak bepalender zijn dan fysieke duels of speldynamiek. Field goals in American football zijn extra relevant omdat er, net zoals bij de schepcorner, sprake is van een samenwerking tussen twee spelers: de aangever (holder) en de kicker.

Drie studies vormen het referentiekader voor deze analyse. Dohmen (2008) onderzocht meer dan 3600 penalty's uit de Duitse Bundesliga en analyseerde via logistische modellen hoe sociale en situationele druk het fenomeen 'choking' beïnvloedt. Jordet et al. (2007) bestudeerden penaltyreeksen in internationale toernooien zoals het WK, waarbij ze stress, ervaring en vermoeidheid kwantificeerden aan de hand van contextuele en fysieke variabelen. Clark et al. (2013) richtten zich dan weer op field goals in American football, en ontwikkelden een uitgebreid model op basis van 11.896 pogingen, waarin zowel fysieke omstandigheden als spelerkenmerken werden opgenomen. Vanuit dit vertrekpunt wordt per thema onderzocht welke factoren terugkeren in deze studies en hoe deze zich verhouden tot de schepcorner.

2.1.1 Druk als bepalende factor

Dohmen (2008) toonde aan dat sociale druk, in het bijzonder het spelen in een thuisstadion, het risico op falen verhoogt bij penalty's in de Bundesliga. Hij spreekt van 'choking' wanneer een speler zonder tussenkomst van de doelman mist. Jordet et al. (2007), die zich richtten op penaltyreeksen in internationale toernooien, bevestigen dat druk een centrale rol speelt: het slagingspercentage daalt naarmate de reeks vordert en is lager in het WK dan in andere toernooien. Clark et al. (2013), die een model bouwden voor field goals in de NFL, vonden daarentegen geen effect van druk op de prestaties van ervaren spelers. Samen suggereren deze studies dat druk vooral impact heeft bij minder ervaren spelers of in sociaal geladen situaties, zoals spelen voor eigen publiek of op het grootste podium.

2.1.2 De rol van ervaring en vermoeidheid

Ervaring wordt in alle drie de studies beschreven en blijkt consequent een beschermende factor tegen falen. Dohmen (2008) operationaliseert dit via het aantal eerder genomen penalty's; Clark et al. (2013) gebruikt het aantal actieve seizoenen. Jordet et al. (2007) benadert ervaring via leeftijd en positie, maar vond geen significante verschillen. Wat vermoeidheid betreft, is alleen Jordet et al. (2007) expliciet: hij gebruikt speelminuten als proxy, maar vindt geen significante impact. Wel stelt hij dat de beperkte vervangingsmogelijkheden de interpretatie bemoeilijken. Vermoeidheid lijkt dus weinig voorspellend bij dit type gestandaardiseerde acties.

2.1.3 Technische vaardigheid als verklarende factor

Ook technische vaardigheid blijkt in meerdere studies een factor van belang, al verschilt de invulling ervan. In de studie van Jordet et al. (2007) wordt vaardigheid benaderd via indirecte proxies zoals de leeftijd van de speler en diens formele positie (aanvaller vs. verdediger), onder de aanname dat aanvallers doorgaans betere afwerkers zijn. Hoewel deze verschillen niet statistisch significant waren, wijzen de auteurs erop dat ervaren aanvallers vaker op cruciale momenten worden ingezet. Dohmen (2008) maakt minder expliciet gebruik van vaardigheid als aparte variabele, maar suggereert dat ervaren spelers consistente prestaties leveren, wat deels vaardigheid kan reflecteren. Clark et al. (2013) nemen geen specifieke vaardigheidsmaat op, maar evalueren kickers over meerdere jaren, wat impliceert dat consistent presterende kickers gemiddeld vaardiger zijn. Toch blijft in alle drie de studies het meten van vaardigheid een methodologisch zwak punt, vaak beperkt tot indirecte benaderingen. In dit onderzoek wordt vaardigheid specifiek gedefinieerd via concrete variabelen, die de kwaliteit van het scheppen aanduiden.

2.1.4 Methodologische benaderingen

Alle drie studies maken gebruik van logistische regressiemodellen om succes te voorspellen. Clark et al. (2013) integreren als enigen omgevingsfactoren zoals wind en temperatuur, wat methodologisch relevant is voor outdoor sporten, maar niet voor een indoorcontext zoals minivoetbal. Dohmen (2008)

hanteert een bijzonder strikte definitie van choking: het missen van een penalty zonder tussenkomst van de doelman, zoals naast of over schieten. Net die afwezigheid van tegenstander maakt zijn analyse zeer relevant voor de schepcorner, waar zonder fysieke tegenstander wordt afgewerkt. Geen van de studies maakt gebruik van machine learning-methoden of modellering van niet-lineaire interacties tussen variabelen.

2.1.5 Samenvattende tabel

Op basis van de besproken studies is in onderstaande tabel weergegeven welke determinanten en modeltypes eerder onderzocht zijn in de context van beslissende spelhervattingen. De variabelen zijn onderverdeeld in psychologische (druk), contextuele (ervaring, vermoeidheid), en technische (vaardigheid) factoren. Daarnaast wordt aangegeven welke voorspellende modellen werden toegepast.

Onderzoek	Sport, spelhervatting	Druk	Ervaring	Vermoeidheid	Vaardigheid	Logistisch model	Random Forest	XGBoost
Dohmen (2008)	Voetbal, penalty	x	x			x		
Jordet et al. (2007)	Voetbal, penalty	x	x	x	x	x		
Clark et al. (2013)	American football, Field goal	x	x			x		
Mijn studie	Minivoetbal, schepcorner	x	x	x	x	x	x	x

Tabel 2: Samenvattende tabel

Wat opvalt is dat eerder onderzoek zich doorgaans beperkt tot een selectie van deze factoren en voornamelijk gebruikmaakt van logistische regressie. In mijn studie wordt daarentegen een holistische benadering gehanteerd waarin al deze determinanten tegelijk worden meegenomen in de analyse van schepcorners. Naast het gebruik van logistische regressie worden ook non-parametrische machine learning methoden zoals Random Forest en XGBoost geïntroduceerd. Deze keuze is methodologisch onderbouwd: dergelijke modellen maken geen strikte aannames over de vorm van relaties tussen variabelen (Breiman, 2001; Song & Lu, 2015), waardoor ze bijzonder geschikt zijn voor het modelleren van de complexe, niet-lineaire interacties. Interactietermen hoeven hierbij niet expliciet gespecificeerd te worden, aangezien de boomstructuur deze automatisch detecteert en integreert tijdens het splitsingsproces. Bovendien tonen meerdere studies aan dat Random Forest en XGBoost superieure voorspellende prestaties leveren ten opzichte van klassieke regressiemodellen (Couronné et al., 2018; Urbanski, 2016). Verder zijn deze modellen robuust tegen overfitting dankzij technieken zoals bootstrapping en random feature selection bij Random Forest, en via structurele elementen zoals subsampling en learning rate bij XGBoost. Ze maken het ook mogelijk om de relatieve bijdrage van elke variabele te interpreteren via zogeheten feature importance-scores, zonder het model te hertrainen. Dit is bij logistische regressie minder evident, onder meer vanwege mogelijke multicollineariteit tussen predictoren. De inzet van deze modellen draagt aldus niet enkel bij aan de voorspelling van succes bij schepcorners, maar ook aan het verkrijgen van diepere inzichten in de onderliggende mechanismen die de uitkomst van deze beslissende spelhervatting bepalen.

Hoewel deze literatuurstudie zich richt op spelhervattingen uit het voetbal en American football, zijn ook in andere sporten gestandaardiseerde acties onderzocht via data-analyse. Zo vertoont baseball op het eerste gezicht sterke parallellen met de schepcorner, denk aan individueel uitvoerbare acties met een duidelijke uitkomst, maar sluiten de meeste toepassingen in de literatuur slechts beperkt aan bij de centrale onderzoeksvraag van deze studie. Koseler en Stephan (2017) voerden een systematisch literatuuroverzicht uit naar machine learning in baseball en categoriseerden de studies in drie hoofdgroepen: regressie, binaire classificatie en multiclass-classificatie. De modellen voor binaire classificatie richten zich echter voornamelijk op het voorspellen van pitch types (bijvoorbeeld fastball vs. non-fastball), of op beslissingen zoals selectie voor een nationaal team of hall-of-fame-opname, waarbij geen sprake is van een scoringssituatie vergelijkbaar met de schepcorner.

Daarnaast richten studies over batting performance zich doorgaans op regressiemodellen. Deze modellen meten continue prestaties over langere periodes, zoals het batting average of het aantal homeruns. Hoewel dit op het eerste gezicht vergelijkbaar lijkt met de schepcorner, sluit het conceptueel minder goed aan. De schepcorner is immers een binair moment, waarbij een actie slaagt of faalt binnen een gestandaardiseerde en stilstaande fase. Omdat deze toepassingen methodologisch wel relevant zijn, maar inhoudelijk onvoldoende vergelijkbaar zijn met de schepcorner, worden ze hier enkel beknopt vermeld als methodologische inspiratiebron, en niet verder inhoudelijk uitgewerkt.

2.2 Onderzoek in nichesporten

Hoewel sportanalytics een snelgroeiend onderzoeksdomein is, blijkt uit een eenvoudige Google Scholar-analyse dat de academische aandacht grotendeels geconcentreerd blijft rond een beperkt aantal populaire sporten. Zo leveren zoekopdrachten als 'soccer analytics' (34.800 hits), 'basketball analytics' (30.000) en 'cycling analytics' (55.300) duizenden resultaten op, terwijl meer niche disciplines zoals 'padel analytics' (1.000) of 'pickleball analytics' (256) beduidend minder zijn vertegenwoordigd. Hoewel deze zoekvolumes geen directe maat zijn voor de wetenschappelijke kwaliteit of diepgang van het onderzoek, suggereren ze wel een duidelijke asymmetrie in onderzoeks aandacht.

Die focus op dominante sporten is tot op zekere hoogte begrijpelijk: populaire sporten tellen wereldwijd meer beoefenaars, beschikken over rijkere datasets en genieten institutionele ondersteuning. Zoals ook blijkt uit het werk van Hulteen et al. (2017) waarin sportparticipatie wereldwijd per sport en leeftijdscategorie in kaart wordt gebracht, beoefent het grootste deel van de bevolking de reeds fervent onderzochte sporten.

Toch leidt het enkel onderzoeken van mainstream sporten tot een blinde vlek. Hoewel Greenhalgh et al. (2011) focust op sportconsumptie bij lacrosse en Crick en Crick (2021) op clubstructuren, wijzen beiden op een toename van de participatie in nichesporten binnen hun domein. Daarnaast duidt grijze literatuur⁴ op een toenemende commerciële en culturele aandacht voor opkomende sporten zoals pickleball, padel en 3x3 basketbal. Hoewel deze bronnen niet peer-reviewed zijn, ondersteunen ze de stelling dat nichesporten als cluster een aanzienlijk en groeiend onderzoeksobject vormen.

⁴ <https://sccgmanagement.com/fr/sccg-articles/2024/11/1/betting-on-the-rise-niche-sports-set-to-explode-in-2025/>

<https://pitchbook.com/news/articles/from-padel-to-volleyball-pe-targets-increasingly-niche-sports>

<https://www.fwwkly.com/2025/01/24/the-rise-of-niche-sports-what-bull-riding-and-pickleball-teach-us-about-fans/>

<https://profluence.com/rise-of-niche-sports/>

Ondanks het beperkte aantal deelnemers per individuele nichesport ontstaat er door de groei van de bestaande nichesporten enerzijds en het ontstaan van nieuwe disciplines anderzijds, een steeds bredere waaier aan onderbelichte sporten. Zo kende het aantal geregistreerde padelspelers sinds 2012 een groei van 175% in Spanje⁵. In de Verenigde Staten verviervoudigde het ledenaantal zelfs in slechts vier jaar tijd, tussen 2020 en 2024⁶. Om aan te tonen dat het aantal nichesporten stijgt, wordt het aantal erkende sportfederaties in Vlaanderen gebruikt als proxy. Sinds 2020 zijn er vijf erkende sportfederaties bijgekomen in Vlaanderen⁷. Deze diversificatie onderstreept de nood aan aangepaste methodologische benaderingen voor het onderzoeken van zulke sporten, zeker wanneer zij kenmerken vertonen die fundamenteel verschillen van de gevestigde, grootschalige sportdisciplines.

Een van deze fundamentele verschillen is de beschikbaarheid van data. In tegenstelling tot populaire sporten, waarvoor vaak reeds uitgebreide statistieken en databanken beschikbaar zijn, ontbreekt dergelijke infrastructuur doorgaans voor minder institutioneel verankerde disciplines. Hierdoor is de onderzoeker aangewezen om zelf data te verzamelen. Dit vereist een doordachte aanpak van zowel de manier waarop data verzameld wordt als het design van de enquêtes.

2.2.1 Vergelijking tussen papieren en online enquêtes

Bij het opzetten van een surveyonderzoek rijst de vraag via welk medium de data best verzameld wordt: via een papieren of een online vragenlijst. Fricker en Schonlau (2002) behoren tot de eerste auteurs die een systematische vergelijking maken tussen deze twee methoden. Zij wijzen erop dat online enquêtes aanzienlijke voordelen bieden op vlak van kostenefficiëntie: printkosten, verzendkosten en manuele gegevensinvoer vallen grotendeels weg. Bovendien worden data onmiddellijk digitaal opgeslagen, wat het risico op fouten verlaagt en de verwerking versnelt. Een bijkomend voordeel is dat online enquêteplatforms technische functionaliteiten bieden zoals inputvalidatie, logische vraagvertakkingen (branching) en doorvraagstructuren, die zowel de datakwaliteit verhogen als de gebruikservaring voor respondenten verbeteren. Ook de snelheid van verspreiding en beantwoording is groter: online vragenlijsten kunnen quasi onmiddellijk verzonden, ingevuld en verwerkt worden. Verder maakt de digitale context het mogelijk om multimedia (zoals afbeeldingen, audio- of videofragmenten) te integreren in de enquête, wat de interactiviteit en betrokkenheid kan versterken.

Verschillende auteurs bevestigen deze voordelen. Wright (2005) wijst, naast de reeds genoemde aspecten, op het bestaan van gebruiksvriendelijke surveysoftware die de technische drempel voor onderzoekers verlaagt. Hij benadrukt eveneens dat online surveys een bijzonder efficiënt instrument zijn om moeilijk bereikbare of geografisch verspreide populaties aan te spreken. Siva Durga Prasad Nayak en Narayan (2019) voegen hieraan toe dat ook opslag- en visualisatiekosten bijzonder laag zijn in een online context. Bovendien benadrukken zij de mogelijkheid tot samenwerking en decentralisatie: onderzoekers kunnen gelijktijdig data verzamelen, verwerken en delen, al vereist dit uiteraard waakzaamheid op vlak van databaseveiliging en privacy. Ball (2019) ten slotte benoemt een bijkomend voordeel van online surveys: het verminderde risico op sociaal wenselijke antwoorden, aangezien er geen interviewer aanwezig is die de respondent mogelijk beïnvloedt. Aangezien dit onderzoek echter uitsluitend werkt met zelfrapportage, zowel voor de online als papieren versie, is dit onderscheid in deze context niet relevant.

⁵ <https://www.padelfip.com/2025/02/padel-in-spain-growth-in-courts-and-player-licenses-in-2024-morcillo-these-results-reflect-the-hard-work-of-recent-years/>

⁶ <https://thepadelstate.com/u-s-padel-association-unveils-first-annual-report/>

⁷ <https://www.sport.vlaanderen/sportfederaties/overzichten-en-cijfers/dossiers-sportfederaties/>

Toch zijn er ook duidelijke nadelen verbonden aan online surveys. Een eerste terugkerend punt in de literatuur betreft de lage en inconsistente responsratio's. Fricker en Schonlau (2002) wijzen erop dat online surveys gemiddeld een lagere deelnamegraad kennen dan hun papieren tegenhangers, al moet deze bevinding genuanceerd worden doordat hun studie dateert uit 2002, toen digitale toegankelijkheid nog beperkter was. Ook Ball (2019) en Wright (2005) wijzen op dit probleem en voegen daaraan toe dat zelfselectie het risico op steekproefvertekening verhoogt: bepaalde personen zijn systematisch sneller geneigd om online deel te nemen, wat de representativiteit van de data ondermijnt. Verder benoemt Ball (2019) dat digitale surveys minder toegankelijk zijn voor respondenten met beperkte digitale vaardigheden, wat zeker relevant blijft ondanks de toegenomen digitalisering. Andere terugkerende beperkingen zijn de verhoogde technische complexiteit bij het opzetten van de enquête, beperkingen in controle over wie deelneemt en het risico op overgeslagen of haastig ingevulde vragen.

Siva Durga Prasad Nayak en Narayan (2019) besteden daarnaast aandacht aan ethische risico's zoals dataprivacy en de moeilijkheden bij het verkrijgen van geïnformeerde toestemming. Ook wijzen zij op het risico van ongewenste hergebruik van data wanneer meerdere onderzoekers samenwerken in een gedeeld platform. Tot slot benadrukt Ball (2019) het belang van een doordachte opbouw van de enquête. Volgens haar zijn een duidelijke introductie, het gebruik van screeningsvragen en helder geformuleerde, goed interpreteerbare items de minimale voorwaarden voor een betrouwbaar online surveyinstrument.

In het licht van deze bevindingen is het belangrijk te onderstrepen dat online surveys bij uitstek geschikt zijn voor cross-sectioneel onderzoek Siva Durga Prasad Nayak en Narayan (2019). Ze bieden de mogelijkheid om efficiënt een ruime groep respondenten te bereiken, data snel te verzamelen en te verwerken en het survey-instrument flexibel af te stemmen op de onderzoeksvragen van deze masterproef. Mits voldoende aandacht voor datakwaliteit, doelgroepgeschiktheid en ethische voorwaarden, vormt de online survey dan ook een gepaste en onderbouwde keuze voor het bestuderen de schepcorner in minivoetbal.

2.2.2 Survey Design

Om betrouwbare en bruikbare inzichten te verkrijgen uit surveyonderzoek is niet alleen de keuze van het medium (papier of digitaal) belangrijk, maar ook het onderliggende ontwerp van de vragenlijst zelf. Hierna wordt een overzicht geboden van best practices rond surveydesign, gebaseerd op toonaangevende literatuur uit de algemene survey-methodologie en uit sport- en gedragswetenschappelijk onderzoek.

Een doordacht surveydesign begint bij een helder geformuleerde onderzoeksvraag, vertaald in concrete en eenduidige items. Verschillende auteurs benadrukken dat surveyvragen niet ad hoc mogen worden opgesteld, maar idealiter voortvloeien uit bestaande theorieën of eerder gevalideerde meetinstrumenten (Draugalis et al., 2008; Krosnick, 2017; Turocy, 2002). In contexten waar dergelijke instrumenten ontbreken, zoals in het onderzoek naar schepcorners in minivoetbal, moet men dit opvangen door systematische vraagontwikkeling, waaronder het pretesten van items en het aftoetsen van inhoudelijke validiteit bij de doelgroep (Ball, 2019; Turocy, 2002).

Wat de formulering van individuele vragen betreft, bestaat er brede consensus over het belang van duidelijke, eenvoudige bewoordingen, het vermijden van samengestelde of leidende vragen, en het gebruik van schaalconstructies met vijf tot zeven antwoordopties (Fanning, 2005; Krosnick, 2017). Antwoordcategorieën moeten bovendien zowel uitputtend als onderling exclusief zijn, om ambiguïteit en overlap te vermijden (Draugalis et al., 2008; Krosnick, 2017; Turocy, 2002).

Voor papieren vragenlijsten, zoals in dit onderzoek gebruikt, wijzen Fanning (2005) en Turocy (2002) op het belang van visuele opmaak: witruimte, consistente lay-out, afgebakende secties en intuïtieve routing zijn essentieel om navigatiefouten te vermijden. Hierin verschillen papieren surveys wezenlijk van digitale formulieren, die via software automatisch branching en validaties kunnen toepassen (Ball, 2019).

Tenslotte benadrukken alle auteurs het belang van pretesten. Niet alleen om technische fouten te vermijden, maar ook om te controleren of respondenten de vragen interpreteren zoals bedoeld. Pretesten geldt dan ook als een onmisbare stap in het waarborgen van content- en face-validiteit (Draugalis et al., 2008; Fanning, 2005; Krosnick, 2017; Turocy, 2002).

Uit deze literatuurstudie blijkt dat de kwaliteit van een survey in grote mate afhangt van de zorgvuldigheid waarmee vragen worden opgesteld, gestructureerd, geformuleerd en getest. Hoewel elk van de besproken auteurs andere accenten legt, bestaat er een duidelijke consensus over de centrale principes van goed surveydesign: helderheid, eenduidigheid, inhoudelijke dekking, validiteit en gebruiksvriendelijkheid. Door deze inzichten te vertalen naar nichesporten zoals het minivoetbal, biedt deze masterproef niet enkel een praktisch onderzoeksinstrument, maar ook een methodologische bijdrage aan het bredere debat over dataverzameling in onderbelichte sportdisciplines.

2.3 Conclusie literatuurstudie

Er bestaat academisch onderzoek naar gestandaardiseerde spelmomenten in populaire sporten die vergelijkbaar zijn met de schepcorner. Zo werd onderzocht welke factoren de uitkomst van een penalty in het veldvoetbal bepalen (Dohmen, 2008; Jordet et al., 2007) en welke factoren invloed hebben op het scoren van een field goal in het American Football (Clark et al., 2013). De meest onderzochte variabelen zijn indicatoren voor druk, ervaring en vermoeidheid. Deze studies maakten gebruik van logistische regressiemodellen om te bepalen welke factoren significant waren voor het succes van het spelmoment. Aangezien er nog geen onderzoek bestaat naar gestandaardiseerde spelsituaties in nichesporten zoals minivoetbal, en er tot nu toe enkel eenvoudige predictiemodellen zijn toegepast, levert deze masterproef een aanvullende bijdrage aan de academische literatuur.

Daarnaast blijkt uit recente literatuur dat nichesporten aan populariteit winnen. Omdat deze sporten vaak weinig institutionele ondersteuning hebben en doorgaans geen centrale databases kennen, is men voor onderzoek aangewezen op zelf ontwikkelde datacollectie. Om die reden werd in deze literatuurstudie ook aandacht besteed aan het gebruik van surveys als dataverzamelingsmethode. Daarbij werd enerzijds gekeken naar het medium van de afname (online versus papier) (Ball, 2019; Fricker & Schonlau, 2002; Siva Durga Prasad Nayak & Narayan, 2019; Wright, 2005) en anderzijds naar het ontwerp van de vragenlijst zelf (Draugalis et al., 2008; Fanning, 2005; Krosnick, 2017; Turocy, 2002). Uit de literatuur blijkt dat het gebruik van online surveys zowel voordelen als nadelen met zich meebrengt. Verder bestaat er veel kennis over algemeen enquêtedesign, maar is er relatief weinig academische literatuur over het ontwerpen van enquêtes specifiek voor sportonderzoek. Deze masterproef levert dan ook niet alleen inhoudelijke inzichten op over minivoetbal, maar draagt ook methodologisch bij aan de manier waarop dataverzameling in nichesporten kan worden opgezet.

Dit onderzoek heeft twee doelstellingen. Enerzijds wil het een volledig beeld geven van de schepcorner in het minivoetbal door de volgende onderzoeksvragen te beantwoorden:

- Hoe bepalend is de schepcorner in het minivoetbal?
- Creëert de schepcorner momentum?
- Wat is het effect van een wijziging van de toekenningsregel (4 corners per schepcorner)?
- Welke factoren beïnvloeden het succes van een schepcorner en welke uitvoeringswijze is het meest effectief?

Anderzijds beoogt deze studie een methodologie te ontwikkelen voor dataverzameling binnen het onderzoek naar nichesporten. Daarbij worden de volgende onderzoeksvragen behandeld:

- Hoe stel je een enquête op voor onderzoek naar een nichesport en welke ontwerpprincipes dienen daarbij in acht te worden genomen?
- Is het beter om een enquête online of op papier af te nemen bij onderzoek naar een nichesport?

3 Methodologie

In de afwezigheid van bestaande centrale datasets over minivoetbal werd in dit onderzoek eigen data verzameld via drie methoden: een online en papieren vragenlijst, video-observaties en directe observatie als kwaliteitscontrole. De vragenlijst werd ontwikkeld op basis van relevante literatuur en diepte-interviews met spelers en experts, en vervolgens gevalideerd via een pretest. De papieren versie werd verspreid in twee sporthallen en ter plaatse ingevuld door wisselers tijdens wedstrijden. Voor wedstrijden op nationaal niveau werd data verzameld via publieke samenvattingen op YouTube.

In totaal werden 603 schepcorners geregistreerd over 231 wedstrijden. De verzamelde ruwe data werd opgeschoond, gestandaardiseerd en uitgebreid via feature engineering. Categorische variabelen werden getransformeerd via one-hot encoding; numerieke variabelen ondergingen normalisatie. Voor het trainen van voorspellende modellen (logistische regressie, Random Forest, en XGBoost) werd een chronologische train-test split gehanteerd, met daaropvolgende hyperparameteroptimalisatie via grid search. De modellen werden geëvalueerd op basis van verschillende prestatie-indicatoren: accuracy, precision, recall, F1-score en AUC. Voor feature selection en om inzicht te krijgen in de belangrijkste variabelen werd gebruikgemaakt van cross-validated permutation importance en Boruta, een wrapper-methode gebaseerd op Random Forest die relevante variabelen identificeert via vergelijking met geshuffelde shadow features. Ten slotte werden aanvullende exploratieve analyses uitgevoerd met betrekking tot impact, momentum en hypothetische spelregelwijzigingen.

3.1 Dataverzameling

Gezien het ontbreken van officiële statistische databanken of centrale datasets over minivoetbal, werd voor dit onderzoek gekozen om eigen data te verzamelen. Hiervoor werden drie methodes ingezet: een survey (online en op papier), video-observatie van wedstrijden en directe observatie. In totaal werden zo 603 schepcorners uit 231 wedstrijden gedocumenteerd en opgenomen in de dataset.

3.1.1 Ontwikkeling van de vragenlijst

De vragenlijst werd ontwikkeld op basis van literatuuronderzoek naar vergelijkbare spelsituaties in meer populaire sporten, en op basis van diepte-interviews met spelers en experts die actief zijn op verschillende niveaus binnen het minivoetbal. Deze interviews dienden als exploratief fundament om relevante variabelen te identificeren en de vraagstelling af te stemmen op de belevingswereld van de doelgroep, zoals aanbevolen in literatuur over enquête onderzoek (Artino et al., 2014). Op basis van deze input werd een voorlopige versie van de vragenlijst opgesteld, die vervolgens werd getest op begrijpelijkheid en volledigheid via een pretest bij grotendeels dezelfde groep van spelers en experts.

3.1.2 Online enquête

Aan het einde van het seizoen 2023–2024 werd via Qualtrics een digitale versie van de vragenlijst verspreid onder alle minivoetbalploegen van kern Gent. Deze verspreiding verliep via e-mail, in samenwerking met de competitieverantwoordelijke. Details over het aantal respondenten en de invulqualiteit worden besproken in de sectie [Resultaten](#).

3.1.3 Papieren enquête

Aan het begin van het seizoen 2024–2025 werden twee papieren versies (zie [Appendix](#)) van de vragenlijst verspreid in twee Gentse sporthallen: Sporthal Sint-Jozef en Sporthal Schoonmeersen. De eerste versie had een chronologische structuur: op de voorzijde werd informatie verzameld die vóór de wedstrijd beschikbaar was (zoals datum, competitieniveau, ploegnamen en ervaringscores),

gevolgd door een invulruimte voor maximaal zes schepcorners. Op de achterzijde werd gevraagd naar gegevens die pas na de wedstrijd beschikbaar waren, zoals de eindscore en het aantal corners volgens het scorebord. De tweede versie verschilde slechts licht en bevatte de vragen over de eindsituatie reeds op de voorzijde.

De formulieren werden door de zaalwachters op de wedstrijdtafel gelegd, waar bankzitters ook de wedstrijdscore bijhielden. Hierdoor konden spelers de vragenlijst onmiddellijk tijdens de wedstrijd invullen. Deze real-time invulling had als doel de datakwaliteit en nauwkeurigheid van de gegevens te verhogen. Onderzoek toont namelijk aan dat real-time dataverzameling betrouwbaardere resultaten oplevert dan retrospectieve bevraging, doordat geheugenvertekening (recall bias) wordt beperkt (Dunton, 2017; Shiffman et al., 2008). De ingevulde formulieren werden nadien manueel ingevoerd in Excel.

3.1.4 Video-observatie van wedstrijden

Voor wedstrijden op het hoogste nationale niveau werd data verzameld via videomateriaal van het YouTube-kanaal '[Kantine 11](#)'⁸. Volgens de beheerder van het kanaal bevatten deze samenvattingen alle schepcorners uit de opgenomen wedstrijden.

3.1.5 Controle van datakwaliteit

Om de betrouwbaarheid van de zelfgerapporteerde gegevens uit de papieren enquêtes te beoordelen, werd een steekproefcontrole uitgevoerd op 20 wedstrijden, resulterend in 54 afzonderlijk geobserveerde schepcorners. Tijdens deze wedstrijden vulde de onderzoeker zelf ter plaatse de vragenlijsten in op basis van observatie, onafhankelijk van de originele zelfrapportages die door spelers waren ingevuld. De overeenkomst tussen beide versies werd geëvalueerd aan de hand van een eenvoudige nauwkeurigheidsmaat (accuracy) en Cohen's kappa, een statistische maat die de interbeoordelaarsovereenstemming corrigeert voor toeval (McHugh, 2012).

Deze controle bood inzicht in de consistentie van de gegevens die via real-time zelfrapportage werden verzameld en leverde een indicatie van de validiteit van deze dataverzamelingsmethode. De resultaten toonden een hoge mate van overeenstemming tussen de zelfrapportage en de observatiegegevens. Voor een gedetailleerde bespreking van deze analyse wordt verwezen naar de sectie [Resultaten](#). In de uiteindelijke dataset werden de zelfgerapporteerde waarden aangepast op basis van deze controle: waar discrepanties werden vastgesteld, werd de gecontroleerde (geobserveerde) waarde overgenomen.

⁸ <https://www.youtube.com/channel/UCFVyKNwmSQ2Flh7D9jqvFAg>

3.1.6 Parameters uit enquête

In totaal werden zo 603 schepcorners uit 231 wedstrijden opgenomen in twee datasets. Hieronder vindt u de opgenomen parameters uit de enquêtes.

Naam	Omschrijving	Datatype
wedstrijd_id	Unieke identificatiecode voor een wedstrijd	Numeriek
datum	Datum waarop de wedstrijd werd gespeeld	Datum
niveau	Competitieniveau van de ploegen in deze wedstrijd	String
uitslag	Einduitslag van de wedstrijd	Vorm: thuis v uit
corners_uitslag	Aantal cornerpunten (niet schepcorner) op het einde van de wedstrijd	Vorm: thuis v uit
klassement_uitslag	Klassementspositie van beide teams voorafgaand aan de wedstrijd	Vorm: thuis v uit

Tabel 3: Wedstrijdparameters

schepcorner_id	Unieke identificatiecode van de schepcorner	Numeriek
wedstrijd_id	Unieke identificatiecode van de wedstrijd waarin de schepcorner plaatsvond	Numeriek
ploegnaam	Naam van de ploeg die de schepcorner kreeg	String
ervaring_schepper	Zelfgerapporteerde ervaring van de schepper	Numeriek (1-5)
ervaring_kopper	Zelfgerapporteerde ervaring van de kopper	Numeriek (1-5)
ervaring_duo	Zelfgerapporteerde ervaring van het schepcornerduo tezamen	Numeriek (1-5)
is_thuisploeg	1 als de ploeg van de schepcorner de thuisploeg is	Binair
kwart	kwart waarin de schepcorner plaatsvond	Numeriek (1-4)
is_schepper_bank	1 als de schepper op de bank zat voor de schepcorner genomen werd	Binair
is_kopper_bank	1 als de kopper op de bank zat voor de schepcorner genomen werd	Binair
score_voor	De score voor de schepcorner genomen word	Binair
schepper_pos_rechts	1 als de schepper rechts van de goal stond vanuit het zicht van de kopper	Binair
positie	Positie van de kopper voor de schepper schept	Numeriek (1-6)
kop_verplaatsen_achter	1 als de kopper naar achter moest verplaatsen om te kunnen koppen	Binair
kop_veel_verplaatsen	1 als de kopper veel moest verplaatsen om te kunnen koppen	Binair
kop_hor_verplaatsen	1 als de kopper naar links of rechts moest verplaatsen om te kunnen koppen	Binair
goal	1 als de schepcorner gescoord werd	Binair

Tabel 4: Schepcornerparameters

3.2 Data preprocessing

3.2.1 Data cleaning

Aangezien de dataset voornamelijk bestond uit zelfgerapporteerde gegevens en deze handmatig ingevoerd werden, was de kans op menselijke fouten aanzienlijk. Een grondige opschoning was dan ook noodzakelijk voordat de data gebruikt kon worden voor analyse.

Een eerste stap bestond uit het standaardiseren van ploegnamen, die vaak op inconsistente wijze waren ingevuld. Zo verwezen 'FC Spitbulls' en 'Spitbulls' naar hetzelfde team. Om dergelijke inconsistenties te detecteren, werd een [fuzzy](#)⁹ matching-algoritme toegepast. De resultaten hiervan werden manueel gecontroleerd en waar nodig geharmoniseerd, zodat elk team over de volledige dataset slechts één unieke naam had.

Daarnaast werd gecontroleerd of het scoreverloop binnen wedstrijden logisch was. De eindscore mocht bijvoorbeeld niet lager zijn dan de tussenscore bij een schepcorner, en een later wedstrijdmoment mocht geen lagere score vertonen dan een eerder moment. Een algoritme werd ontwikkeld om deze inconsistenties op te sporen, waarna verdachte observaties manueel werden gevalideerd. Veel van de geconstateerde fouten bleken het gevolg van invoerfouten in Excel of verwisseling van thuis- en uitploeg in de enquête.

Verder werd de geldigheidsrange per kolom gecontroleerd. Indien waarden niet aan de verwachte input voldeden, werd teruggегреpen naar de originele papieren enquête om de correcte waarde in te voeren. Missende einduitslagen, speeldatum en niveaus werden geïmputeerd via een externe bron, de website van [Voetbal Vlaanderen](#)¹⁰.

Ontbrekende waarden betreffende de ervaring van spelers werden eerst aangevuld met eerder ingevulde teamdata. Indien deze ontbraken, werd gebruikgemaakt van het gemiddelde ervaringsniveau binnen het betrokken niveau. Missende waarden in klassementsinformatie werden geïmputeerd op basis van eerder opgegeven rangschikkingen bij andere wedstrijden. Indien die info ontbrak, werden de klassement variabelen geïmputeerd door gemiddelden.

3.2.2 Transformaties

Alle categorische variabelen werden getransformeerd via One-Hot Encoding, een gangbare techniek die goed werkt in combinatie met lineaire modellen (Segar, 2018). De numerieke variabelen werden genormaliseerd met behulp van een StandardScaler, een methode die breed ondersteund wordt in de literatuur om schaalverschillen tussen variabelen te reduceren (Ahsan et al., 2021; de Amorim et al., 2023).

⁹ <https://pypi.org/project/RapidFuzz/>

¹⁰ <https://www.voetbalvlaanderen.be/>

3.2.3 Feature engineering

Om het voorspellend vermogen van de modellen te vergroten en de context van schepcorners accurater te vatten, werd uitgebreide feature engineering toegepast. Naast de rechtstreeks uit de observatieformulieren overgenomen variabelen, werden aanvullende kenmerken gecreëerd.

Naam	Omschrijving	Datatype
score_vershil_voor	Scoreverschil tussen de ploeg met de schepcorner en de tegenstander voor het nemen van de schepcorner	Numeriek
klassement	Positie van de ploeg in het klassement	Numeriek
klassement_verschil	Verschil in klassementspositie tussen beide ploegen	Numeriek
is_degradatieplaats	1 als de ploeg op een degradatieplaats staat	Binair
is_promotieplaats	1 als de ploeg op een promotieplaats staat	Binair
wedstrijd_corner_nr	De hoeveelste schepcorner het is in de wedstrijd	Numeriek
ploeg_wedstrijd_corner_nr	De hoeveelste schepcorner het is voor de ploeg in de wedstrijd	Numeriek
goal_lag_ploeg	1 als de vorige schepcorner door de ploeg werd gescoord	Binair
goal_lag_wedstrijd	1 als de vorige schepcorner in de wedstrijd werd gescoord	Binair
positie_verticaal_voor	1 als de koper vooraan stond opgesteld	Binair
positie_horizontaal	Horizontale positie van de koper (Links, Midden of Rechts)	String
ploeg_scoreperc	Historisch scorepercentage van de ploeg	Numeriek
algemene_ervaring	Gemiddelde van de ervaringsscores van schepper, koper en duo	Numeriek
niveau	De niveaus herverdeeld in drie categorieën: kern_laag, kern_hoog en nationaal	String

Tabel 5: Geconstrueerde variabelen

Bijzondere aandacht ging uit naar het modelleren van de mentale druk bij schepcorners. Deze druk werd benaderd op basis van zowel de bredere wedstrijdcontext als de actuele situatie in de match. De bredere context omvat promotie- of degradatiebelangen, maar ook wedstrijden tussen ploegen die dicht bij elkaar staan in het klassement. De actuele druk werd weergegeven via het scoreverschil op het moment van de schepcorner. Scoreverschillen en klassementspositie fungeren daarbij als

indicatoren voor de relatieve waarde van het moment, wat relevant is voor het inschatten van psychologische druk (Bransen et al., 2019).

Om het effect van recent succes en vertrouwen te modelleren, werden ook variabelen opgenomen die eerdere prestaties bij schepcorners weerspiegelen. Dit omvat onder meer het scoren bij vorige schepcorners en het historisch scorepercentage van het team.

Daarnaast werd de positionele informatie van de koper opgesplitst in een verticale en een horizontale component, om verschillende dimensies van opstelling te onderscheiden.

Verder bleek dat de ervaringsvariabelen sterk met elkaar correleerden. Multicollineariteit leidt in logistische regressie tot onnauwkeurige en instabiele coëfficiëntschattingen, wat de voorspellende waarde van het model aanzienlijk kan ondermijnen (Senaviratna et al., 2019). Hoewel dit bij tree-based modellen doorgaans minder directe gevolgen heeft voor de predictieve prestaties, bemoeilijkt het wel degelijk de interpretatie van de feature importance van variabelen (Gregorutti et al., 2017; Toloşi et al., 2011). Aangezien dit onderzoek in hoofdzaak gericht is op het identificeren van de belangrijkste bepalende factoren achter het succes van een schepcorner, werd ervoor gekozen om de sterk gecorreleerde ervaringsvariabelen samen te voegen tot één geaggregeerde variabele, een aanpak die ook in de literatuur wordt aanbevolen (Senaviratna et al., 2019).

Ten slotte werd besloten om de categorische variabele niveau terug te brengen van zeven afzonderlijke categorieën naar drie overkoepelende klassen. De niveaus werden herverdeeld in: *kern_laag* (kernniveau 3 en 4), *kern_hoog* (kernniveau 1 en 2) en *nationaal* (alle niet-kernniveaus). Deze reductie werd doorgevoerd omdat sommige oorspronkelijke categorieën, zoals nationaal 4 en provinciaal 2, slechts in zeer beperkte mate voorkwamen. Het samenvoegen van dergelijke zeldzame categorieën verhoogt de stabiliteit van het model, voorkomt overfitting en bevordert de generaliseerbaarheid naar ongeziene data .

3.3 Model selection

In dit onderzoek werd logistische regressie gebruikt als basismodel voor het voorspellen van de uitkomst van een schepcorner. Deze keuze werd ingegeven door de brede toepassing van logistische regressie in eerdere literatuur als een eenvoudig, interpreteerbaar en effectief startpunt voor binaire classificatieproblemen. Bovendien werd dit model gebruikt als referentiepunt voor de prestaties van complexere modellen. Aangezien logistische regressie een white-box model is, waarbij de invloed van elke feature rechtstreeks afgeleid kan worden uit de regressiecoëfficiënten, werden deze coëfficiënten gebruikt om het effect van elke predictor op de kans op een succesvolle schepcorner te interpreteren (Peng et al., 2002).

Vervolgens werd een Random Forest-model toegepast. Deze algoritmische keuze is verantwoord omdat Random Forests robuust zijn tegen overfitting, zelfs bij een relatief beperkte dataset, dankzij het bagging-mechanisme dat gebruik maakt van verschillende boomstructuren. Bovendien laat dit model toe om op een intuïtieve manier de feature importance te bepalen op basis van gemiddeld informatieverlies (Breiman, 2001).

Daarnaast werd ook XGBoost ingezet. Dit model werd gekozen omwille van zijn uitstekende prestaties bij gestructureerde data, zijn ingebouwde regularisatie, en zijn hoge robuustheid tegenover overfitting (Chen et al., 2025; Urbanski, 2016). Net zoals bij Random Forests, biedt XGBoost inzicht in feature importance, wat toelaat om modeltransparantie te combineren met hoge prestaties.

De schepkwaliteit-indicatoren zijn variabelen die aangeven of, en in welke richting, de koper zich moet verplaatsen om de bal te kunnen koppen. Het gaat om de variabelen `kop_verplaatsen_achter`,

kop_veel_verplaatsen en kop_hor_verplaatsen. Uit bevraging van experts bleek dat een schep kwaliteitsvol is als de koper niet of slechts een minimale voorwaartse beweging moet maken. Deze variabelen worden daarom geïnterpreteerd als een maat voor schepcornerkwaliteit.

Aangezien deze informatie pas beschikbaar is na de uitvoering van de eerste fase van de schepcorner, zijn de indicatoren in de praktijk niet bruikbaar op het moment van voorspelling. Hun opname in een voorspellend model zou de toepasbaarheid ervan in realistische omstandigheden kunnen beperken. Om dit te onderzoeken werd, naast een eerste analyse mét deze variabelen als voorspellers, ook een alternatieve analyse uitgevoerd waarin de schepkwaliteit-indicatoren werden weggelaten. In beide gevallen werden identieke modelstructuren en trainingsprocedures gehanteerd om een eerlijke vergelijking mogelijk te maken.

3.4 Model training en evaluatie

Voor het trainen van de modellen werd de dataset eerst opgesplitst in een trainings- en testset met het oog op een correcte modelvalidatie en om data leakage te voorkomen. Omdat de gegevens een tijdsdimensie bevatten, werd de dataset eerst gesorteerd op *datum* en *schepcorner_id* zodat de chronologische volgorde binnen wedstrijden gegarandeerd is. Vervolgens werd een chronologische split toegepast, waarbij de eerste 80% van de observaties werd gebruikt als trainingsset en de resterende 20% als testset. Deze aanpak weerspiegelt een realistische voorspelsituatie, waarbij het model leert op basis van vroegere data en getest wordt op latere, ongeziene waarnemingen. Dit is belangrijk aangezien verschillende features zijn afgeleid uit eerdere schepcorners.

Voor het trainen van alle modellen, met uitzondering van het baseline logistische regressiemodel, werd hyperparameteroptimalisatie toegepast. Daarbij werd voor grid search gekozen, in combinatie met cross-validation, om systematisch alle relevante hyperparametercombinaties te evalueren. Deze keuze is gebaseerd op de bevindingen van Yang en Shami (2020), die aantonen dat grid search een robuuste en transparante optimalisatietechniek is, vooral wanneer de hyperparameterspace beperkt en goed afgebakend is. Hoewel random search in grotere en minder gestructureerde zoekruimtes vaak efficiënter is (Bergstra et al., 2012), bleek uit onze concrete configuratie, met relatief weinig hyperparameters en discrete waarden, dat grid search in dit geval volledige dekking en reproduceerbaarheid biedt zonder overmatige rekentijd. De volledige parameterroosters en de optimale hyperparameters voor respectievelijk het Random Forest- en XGBoost-model zijn terug te vinden in de [Appendix](#).

In eerste instantie werd de modelprestatie beoordeeld aan de hand van accuracy. Deze maatstaf is echter gevoelig voor de dominantie van de meerderheidsklasse en kan bij ongelijke klassenverdelingen een vertekend beeld geven van de werkelijke prestaties van het model (Hossin & Sulaiman, 2015). Bovendien is het in het algemeen niet aangewezen om de modelkwaliteit uitsluitend op basis van accuracy te beoordelen (Demšar, 2006). Om die reden werd ervoor gekozen om aanvullende evaluatiemaatstaven te gebruiken die een genuanceerder beeld bieden.

Precision werd ingezet om te meten welk aandeel van de voorspellingen voor de positieve klasse effectief correct was. Recall werd berekend op de minderheidsklasse, namelijk het missen van een schepcorner, en geeft weer in welke mate het model erin slaagt om gemiste kansen correct te identificeren. De F1-score combineert precision en recall in één harmonisch gemiddelde en vormt zo een gebalanceerde maatstaf wanneer beide fouttypes, fout-positieven en fout-negatieven, even zwaar doorwegen.

Ten slotte werd ook de AUC (Area Under the ROC Curve) gebruikt. Deze maat geeft een betrouwbaar beeld van het onderscheidend vermogen van het model over een volledig bereik van

classificatiedrempels en blijft bovendien robuust in situaties met lichte class imbalance (Bradley, 1997; Fawcett, 2006). Bovendien is in eerder onderzoek aangetoond dat AUC een beter evaluatiecriterium is dan accuracy (Huang & Ling, 2005). Daarom werd deze uiteindelijk gehanteerd als hoofdcriterium voor modelvergelijking. Door het gebruik van deze aanvullende maatstaven ontstaat een breder en accurater beeld van de prestaties van het model dan op basis van alleen accuracy mogelijk zou zijn.

Om de AUC-scores van verschillende modellen op statistisch verantwoorde wijze met elkaar te vergelijken, werd gebruikgemaakt van de DeLong-test (DeLong et al., 1988). Deze methode is geschikt voor het vergelijken van gecorreleerde ROC-curves, zoals het geval is wanneer modellen op dezelfde testset worden geëvalueerd, en laat toe om formeel te bepalen of het verschil in AUC statistisch significant is.

Om irrelevante of mogelijk misleidende variabelen uit het model te filteren, werd gebruikgemaakt van permutation importance zoals geïmplementeerd in de scikit-learn-bibliotheek. Deze techniek meet per feature hoeveel de modelprestatie (AUC) daalt wanneer de waarden van die feature willekeurig worden gepermuteerd. Het idee hierachter is dat belangrijke features leiden tot een duidelijke daling in modelprestatie zodra ze willekeurig aan een observatie worden toegewezen.

Om de betrouwbaarheid van deze inschattingen te verhogen, werd de permutation importance berekend in een cross-validated opzet. Hierbij werd het model meerdere keren getraind en geëvalueerd op verschillende splits van de testset. Deze aanpak levert stabielere resultaten op dan één enkele meting, en voorkomt bovendien dat informatie uit het trainingsproces doorsijpelt in de evaluatie, wat een risico op data leakage zou vormen.

Na het berekenen van de permutation importance werd een filter toegepast waarbij alle features met een negatieve gemiddelde impact werden verwijderd. Dergelijke features kunnen het model zelfs beter laten presteren wanneer ze worden verstoord, wat suggereert dat ze eerder ruis dan informatie bevatten.

Daarnaast werd ook gebruikgemaakt van Boruta als methode voor feature selection. Boruta staat bekend als een conservatieve techniek die ernaar streeft om alle relevante variabelen te behouden. Dit maakt de methode bijzonder geschikt voor dit onderzoek, waarin niet alleen het voorspellen centraal staat, maar ook het begrijpen van welke factoren bijdragen aan het succes van een schepcorner. Voor een uitgebreide beschrijving van het algoritme wordt verwezen naar de ontwikkelaars van het algoritme (Kursa & Rudnicki, 2010). Aangezien de analyses in Python werden uitgevoerd, werd gebruikgemaakt van de specifieke Python-implementatie van het Boruta-algoritme ([BorutaPy-package](https://pypi.org/project/Boruta/)¹¹).

3.5 Exploratieve analyses

Naast de voorspellende modellen werden een reeks exploratieve analyses uitgevoerd, met als doel inzicht te krijgen in de determinanten van succesvolle schepcorners en hun bredere invloed binnen het wedstrijdverloop.

3.5.1 Impactanalyse

Zo werd onderzocht wat de impact van een schepcorner is op het resultaat van een wedstrijd. Eerst werd nagegaan in hoeveel wedstrijden minstens één schepcorner voorkwam. Vervolgens werd onderzocht hoe vaak teams minstens één schepcorner per wedstrijd konden verwachten.

¹¹ <https://pypi.org/project/Boruta/>

Vervolgens werden vormen van impact gedefinieerd. Er is sprake van impact wanneer de uitslag van een wedstrijd zou wijzigen. Een wijziging van de wedstrijduitkomst wordt in deze analyse gedefinieerd als een verandering waarbij het eindresultaat leidt tot een andere winnaar, of tot het omslaan van een overwinning in een gelijkspel, of omgekeerd. Een wijziging van 7–2 naar 5–2 wordt bijvoorbeeld niet beschouwd als impact, aangezien het thuisteam in beide gevallen wint. Enkel wanneer het resultaat een ander competitief eindverloop weerspiegelt, wordt gesproken van een impact op de wedstrijduitkomst.

- **Strikte impact:** hierbij werd nagegaan in hoeveel wedstrijden het resultaat zou wijzigen indien slechts één schepcorner van uitkomst mag veranderen (van gemist naar gescoord of omgekeerd).
- **Milde impact:** in deze variant wordt aangenomen dat alle schepcorners in een wedstrijd hun uitkomst kunnen wijzigen, waarna geanalyseerd wordt hoe vaak dit het wedstrijdresultaat beïnvloedt.
- **Hypothetische extremen:** daarnaast werd nagegaan wat er zou gebeuren in twee hypothetische scenario's: (1) geen enkele schepcorner resulteert in een doelpunt, en (2) elke schepcorner wordt benut. In beide gevallen werd ook de impact op de wedstrijduitslag bepaald.

Deze impactanalyse laat toe om het strategische belang van schepcorners binnen wedstrijden beter te begrijpen en biedt inzicht in hoe kleinschalige spelmomenten het bredere verloop kunnen beïnvloeden.

3.5.2 Momentumanalyse

Ten tweede werd een momentumanalyse uitgevoerd om na te gaan of het nemen van een schepcorner invloed uitoefent op het verdere verloop van de wedstrijd. Deze analyse werd op twee manieren aangepakt:

- **Vergelijking van momentum direct na de schepcorner:** voor elke schepcorner werd het scoringsverloop tussen de schepcorner en de volgende gebeurtenis (volgende corner of einde van de wedstrijd) geanalyseerd. Er werd een label toegekend op basis van de doelpuntenbalans na de schepcorner: positief momentum (meer doelpunten door het team dat de schepcorner nam), neutraal (gelijke score), of negatief (meer doelpunten door de tegenpartij). Deze verdelingen werden vervolgens vergeleken tussen gescoorde en gemiste schepcorners. Omdat specifiek het additionele momentum na een schepcorner interessant is, werd een eventueel doelpunt dat rechtstreeks uit de schepcorner voortkwam niet meegeteld in de scoringsbalans voor de daaropvolgende periode.
- **Vergelijking vóór en na de schepcorner:** hier werd voor elke situatie het momentum vóór de schepcorner vergeleken met het momentum erna. Een schepcorner kreeg een verbetering label indien het momentum er positief op vooruitging (bijv. van negatief naar neutraal of positief), een verslechtering label indien ze achteruitging, en een ongewijzigd label indien er geen verandering was. Opnieuw werden de verdelingen vergeleken.

3.5.3 Effect van een hypothetische regelwijziging

Daarna werd de impact onderzocht van een hypothetische regelwijziging waarbij het aantal corners dat leidt tot een schepcorner werd aangepast van vier naar respectievelijk drie of vijf. In eerste instantie werd berekend hoeveel schepcorners er procentueel extra of minder verwacht zouden

worden onder elk scenario. Dit gebeurde door per wedstrijd het totaal aantal corners per ploeg te bepalen en vervolgens te berekenen hoeveel schepcorners dit zou opleveren bij toepassing van een toekenningregel van drie of vijf corners per schepcorner. Ten slotte werden deze gesimuleerde totalen vergeleken met het werkelijke aantal schepcorners op basis van de huidige regel (één schepcorner per vier corners).

Daarnaast werd nagegaan wat de impact van een regelwijziging zou zijn op ploegniveau. Eerst werd per ploeg het aantal extra of minder gescoorde doelpunten per seizoen gesimuleerd. Dit gebeurde door het gemiddelde aantal schepcorners per wedstrijd te combineren met het bijbehorende scorepercentage, en dit resultaat vervolgens te vermenigvuldigen met het procentuele verschil tussen het verwachte aantal schepcorners onder de nieuwe regel en het huidige aantal. Op die manier werd het verwachte verschil in doelpunten per ploeg per seizoen geschat.

Vervolgens werd onderzocht wat de regelwijziging zou betekenen voor het behaalde puntenaantal en de eindrangschikking per ploeg. Voor elke wedstrijd van de ploeg werd gesimuleerd wat de uitslag zou zijn indien de regelwijziging van kracht was geweest. Vervolgens werd nagegaan of dit tot een andere wedstrijduitkomst (winst, gelijkspel of verlies) zou hebben geleid. In dat geval werd de impact op het aantal competitiepunten bepaald. Deze puntenverschillen werden opgeteld over alle wedstrijden, waarna werd berekend welke invloed dit zou hebben gehad op de eindrangschikking van de ploeg in het klassement. De informatie over de oorspronkelijke eindrangschikking werd daarbij ontleend aan de website van [Voetbal Vlaanderen](https://www.voetbalvlaanderen.be/)¹².

3.5.4 Analyse van variabelen

Tenslotte werd het scorepercentage per categorie van verschillende verklarende variabelen berekend. Dit bood een eerste indicatie van mogelijke verbanden en trends tussen bepaalde kenmerken en het succes van de schepcorner. Om na te gaan of de geobserveerde verschillen ook statistisch significant waren, werden vervolgens geschikte toetsen toegepast.

De chi-kwadraattoets werd ingezet om te onderzoeken of er een significant verschil bestaat tussen geobserveerde en verwachte frequenties in categorische data. Bij vergelijkingen tussen proporties van twee groepen werd gebruikgemaakt van een Z-test voor proportieverschillen, een methode die traditioneel wordt toegepast in vergelijkende steekproefanalyses. Om te toetsen op lineaire trends in proporties over ordinale categorieën (zoals *niveau* en *kwart*), werd de Cochran-Armitage trendtest gebruikt (Armitage, 1955). Deze test laat toe om mogelijke monotone trends in dichotome uitkomsten over geordende groepen op te sporen. Ten slotte werd de Kruskal-Wallis-toets (Kruskal & Wallis, 1952), een niet-parametrische alternatief voor ANOVA, toegepast om te bepalen of het gemiddelde aantal schepcorners per wedstrijd significant verschilt tussen de verschillende competitieniveaus.

Alle analyses en modellen werden uitgevoerd in Python. De gebruikte code is publiek beschikbaar via GitHub en kan geraadpleegd worden op: https://github.com/simon-cosyns/thesis_schepcorner.

¹² <https://www.voetbalvlaanderen.be/>

4 Resultaten

Dit hoofdstuk presenteert de bevindingen van het onderzoek naar de schepcorner in het minivoetbal. De resultaten worden gepresenteerd in functie van de centrale onderzoeksvragen:

1. Welke ontwerpprincipes zijn het meest geschikt voor het ontwikkelen van een enquête over een nichesport?
2. Welke afnamevorm, online of op papier, is het meest geschikt voor dit type onderzoek?
3. Hoe bepalend is de schepcorner in het minivoetbal?
4. Draagt de schepcorner bij aan het creëren van momentum?
5. Hoe beïnvloedt een regelwijziging omtrent schepcorners het totale aantal schepcorners en wat is het effect voor specifieke ploegen?
6. Welke factoren beïnvloeden het succes van een schepcorner en welke uitvoeringswijze is het meest effectief?

In lijn met de opbouw van dit onderzoek worden eerst de resultaten met betrekking tot de dataverzameling en de enquêtecomponent besproken. Hierbij komen onder meer de voltooiingsgraad, de datakwaliteit en de controle op zelfrapportage via observatievergelijking aan bod. Vervolgens worden de analyses besproken met betrekking tot de impact en het momentum van schepcorners, en ten slotte komen de resultaten aan bod van de classificatiemodellen die het succes van een schepcorner trachten te voorspellen op basis van uiteenlopende variabelen.

De interpretatie van deze resultaten, evenals hun betekenis in het bredere sportwetenschappelijke en beleidsmatige kader, wordt behandeld in het daaropvolgende discussiegedeelte.

4.1 Resultaten enquête en dataverzameling

4.1.1 Online enquête

In het kader van de dataverzameling werd aan het einde van het seizoen 2023–2024 eerst een online enquête verspreid. De enquête werd in totaal 37 keer geopend. Daarvan werden 22 enquêtes ingevuld, waarvan 19 volledig en correct. Twaalf van de correct ingevulde formulieren waren afkomstig van de onderzoeker zelf. Van de resterende 25 externe antwoorden, bleken slechts zeven enquêtes zowel volledig als correct ingevuld. Dit komt overeen met een voltooiingsgraad van 28%.

Omdat deze ingevulde enquêtes vaak betrekking hadden op bekerwedstrijden, jeugdwedstrijden of vrouwenteams, die inhoudelijk niet goed aansloten bij de overige observaties uit competitiewedstrijden, en vanwege de bijzonder lage voltooiingsgraad, werd besloten om de online data niet verder mee te nemen in de analyse.

Om deze lage voltooiing te kunnen kaderen, werd nagegaan hoeveel ploegen effectief bereikt werden en hoe de respons zich verhoudt tot gemiddelden uit eerder onderzoek. Hoewel er geen exacte gegevens beschikbaar zijn over het aantal resterende wedstrijden per ploeg, is wel bekend dat alle 287 ploegen die dat seizoen actief waren in kern Gent de vragenlijst ontvingen en op dat moment nog minstens één wedstrijd moesten afwerken. Op basis hiervan kan worden vastgesteld dat de maximale responsgraad slechts 2,4% bedroeg. Dit ligt aanzienlijk lager dan de gemiddelde online responsgraad van 44,1% (Wu et al., 2022). Een belangrijke kanttekening is dat het gemiddelde van 44,1% afkomstig is uit onderwijsgerelateerd onderzoek.

4.1.2 Papieren enquête

De papieren enquête leverde aanzienlijk meer bruikbare gegevens op. In totaal werden 208 papieren formulieren verzameld, waarvan er 19 door de onderzoeker zelf en 189 door externe respondenten werden ingevuld. Hiervoor werden twee versies van het formulier gebruikt. De eerste versie had een chronologische opbouw: algemene wedstrijd informatie stond op de voorzijde, terwijl de eindscore en het aantal cornerpunten pas bij het eindsignaal op de achterzijde werden genoteerd. Deze versie werd 163 keer ingevuld door externe respondenten. De tweede versie, waarbij ook de informatie over het einde van de wedstrijd op de voorzijde stond, werd 26 keer ingevuld door externe respondenten. Ook in dit geval was het moeilijk om de responsgraad te bepalen, aangezien de papieren enquêtes door zaalwachters bij de scoreborden werden geplaatst. Hierdoor is niet bekend hoeveel formulieren daadwerkelijk zijn uitgedeeld.

Bij vergelijking van beide versies bleek de invulgraad van de algemene eindinformatie (eindscore en aantal cornerpunten) duidelijk hoger in versie twee. In de eerste versie werd deze sectie slechts in 102 van de 163 gevallen volledig ingevuld (de enquêtes ingevuld door de onderzoeker werden hier niet meegerekend), terwijl in de tweede versie 25 van de 26 formulieren correct en volledig werden ingevuld. Een proportietoets toonde aan dat versie twee significant vaker volledig werd ingevuld dan versie één ($Z = -3.39$, $p < .001$). Dit sterk significante verschil wijst erop dat de tweede versie een effectievere opmaak had.

4.1.3 Video analyse

Tot slot werden 23 wedstrijden geanalyseerd op basis van videomateriaal, afkomstig van het YouTube-kanaal '[Kantine 11](https://www.youtube.com/channel/UCFVyKNwmSQ2Flh7D9jqvFAg)¹³'. Hoewel deze bron bruikbaar was om schepcorners te observeren, werd de stand van het aantal corners aan het eind van de wedstrijd vaak niet vermeld, wat de volledigheid van deze gegevens beperkt. Een samenvattend overzicht van de drie methoden is weergegeven in de tabel hieronder.

Methode	Externe respons	Respons onderzoeker	Percentage volledig ingevuld (extern)
Online enquête	7	12	28%
Papieren versie 1	163	19	63%
Papieren versie 2	26	0	96%
Videoanalyse	n.v.t	23	n.v.t

Tabel 6: Overzicht dataverzamelmethode

¹³ <https://www.youtube.com/channel/UCFVyKNwmSQ2Flh7D9jqvFAg>

4.1.4 Data controle

De onderzoeker observeerde 20 wedstrijden, waarin 54 schepcorners werden geregistreerd, met als doel de kwaliteit van de ingevulde enquêtes te evalueren. Deze steekproef maakte het mogelijk om de betrouwbaarheid van de door spelers gerapporteerde gegevens te toetsen aan objectieve observaties. Op basis van een vergelijking tussen de observaties en de enquêtegegevens werden zowel de nauwkeurigheid (*accuracy*) als de overeenstemming (*Cohen's kappa*) berekend. Deze statistieken bieden inzicht in de mate waarin respondenten correcte en consistente informatie aanleverden, een essentiële voorwaarde voor het waarborgen van de validiteit van de volledige dataset. De resultaten hiervan zijn weergegeven in onderstaande tabel.

Variabele	Accuracy	Cohen's kappa
niveau	100,00%	1,00
uitslag	100,00%	1,00
corners_uitslag	85,19%	0,83
kwart	98,15%	0,97
is_schepper_bank	98,15%	0,92
is_kopper_bank	100,00%	1,00
score_voor	98,15%	0,98
schepper_pos_rechts	100,00%	1,00
positie	85,19%	0,72
kop_verplaatsen_achter	98,15%	0,88
kop_veel_verplaatsen	87,04%	0,59
kop_hor_verplaatsen	90,74%	0,78
goal	100,00%	1,00

Tabel 7: Overeenstemming controle en zelfgerapporteerde enquêtes

Uit de analyse blijkt dat de overeenstemming tussen observaties en enquêteresultaten over het algemeen zeer hoog is. De accuracy ligt in alle gevallen boven de 85%, en ook Cohen's kappa is voor de meeste variabelen hoog. Volgens de interpretatierichtlijnen van Landis en Koch (1977) wijst een kappa tussen 0,61 en 0,80 op substantiële overeenkomst, terwijl waarden boven 0,81 duiden op bijna perfecte overeenkomst. Niettemin is het belangrijk kritisch te blijven. McHugh (2012) betoogt dat een kappa onder de 0,60 wijst op onvoldoende betrouwbaarheid en dat bij dergelijke waarden het vertrouwen in de consistentie van de gegevens beperkt moet blijven. Indien men deze strengere drempel hanteert, valt de kappa-waarde voor de variabele *kop_veel_verplaatsen* net onder de aanvaarde grens. De relatief lagere kappa-waarde voor deze variabele lijkt echter vooral het gevolg te zijn van een bekende beperking in de berekening van kappa. Cohen's kappa corrigeert namelijk voor toevallige overeenstemming, waarbij de verwachte overeenkomst wordt berekend op basis van de marginale distributies. In het geval van de variabele *kop_veel_verplaatsen* zijn die verdelingen sterk scheef: 41 'nee'-antwoorden tegenover 13 'ja' in de enquête, en 46 tegenover 8 in de controle. Volgens Zec et al. (2017) leidt zo'n scheve verdeling tot de zogeheten kappa-paradox, waarbij de kappa-waarde kunstmatig laag uitvalt ondanks een hoge feitelijke overeenkomst. Toch werd *kop_veel_verplaatsen* in 7 van de 54 gevallen verschillend beoordeeld tussen enquête en observatie, wat meer is dan bij andere variabelen, die naar de verplaatsing van de koper peilen. Mogelijk hangt dit samen met het meetconcept zelf. Deze variabele vereist een inschatting van de afstand die de koper verplaatst, een schaalgevoelige en dus subjectieve beoordeling. Daarentegen zijn de andere gelijkaardige indicatoren (*kop_verplaatsen_achter* en *kop_hor_verplaatsen*) binair en visueel eenvoudiger vast te stellen: ofwel beweegt een speler zich achteruit, ofwel niet. Deze laatste variabelen zijn minder vatbaar voor interpretatieverschillen en werden dan ook vaker consistent beoordeeld.

Daarom kan op basis van deze bevindingen geconcludeerd worden dat er een hoge mate van overeenstemming bestaat tussen de ingevulde enquêtes en de onafhankelijke observaties. Dit wijst erop dat respondenten in staat waren om nauwkeurig en consistent verslag uit te brengen van de schepcorners, wat de betrouwbaarheid van de bredere dataset ondersteunt.

4.2 Impact

Uit de verzamelde dataset van 231 wedstrijden blijkt dat in slechts één wedstrijd geen enkele schepcorner werd uitgevoerd. Dat betekent dat in 99,57% van de wedstrijden minstens één schepcorner voorkwam. Wanneer gekeken wordt naar de verdeling per team, blijkt dat een team in 88,96% van de wedstrijden minstens één schepcorner mag verwachten. In 78,35% van de wedstrijden kregen beide teams minstens één schepcorner, in 21,21% slechts één team, en in 0,43% geen van beide teams.

De impact van schepcorners op het wedstrijdverloop werd geanalyseerd aan de hand van vier scenario's, waarvan de resultaten zijn weergegeven in onderstaande tabel. Een wijziging van de wedstrijduitkomst werd hierbij gedefinieerd zoals uiteengezet in de [Methodologie](#) sectie.

Scenario	Definitie	Aantal wedstrijden	Percentage
Strikte impact	Wedstrijduitkomst verandert bij wijziging van slechts één schepcorner	59	25,54%
Milde impact	Wedstrijduitkomst verandert wanneer alle schepcorners mogen veranderen van uitkomst	139	60,17%
Alle schepcorners gemist (of geen schepcorners meer)	Impact als alle scores gemist worden. Of er geen schepcorners meer zijn	42	18,18%
Alle schepcorners gescoord	Impact als alle schepcorners worden omgezet in een doelpunt	28	12,12%

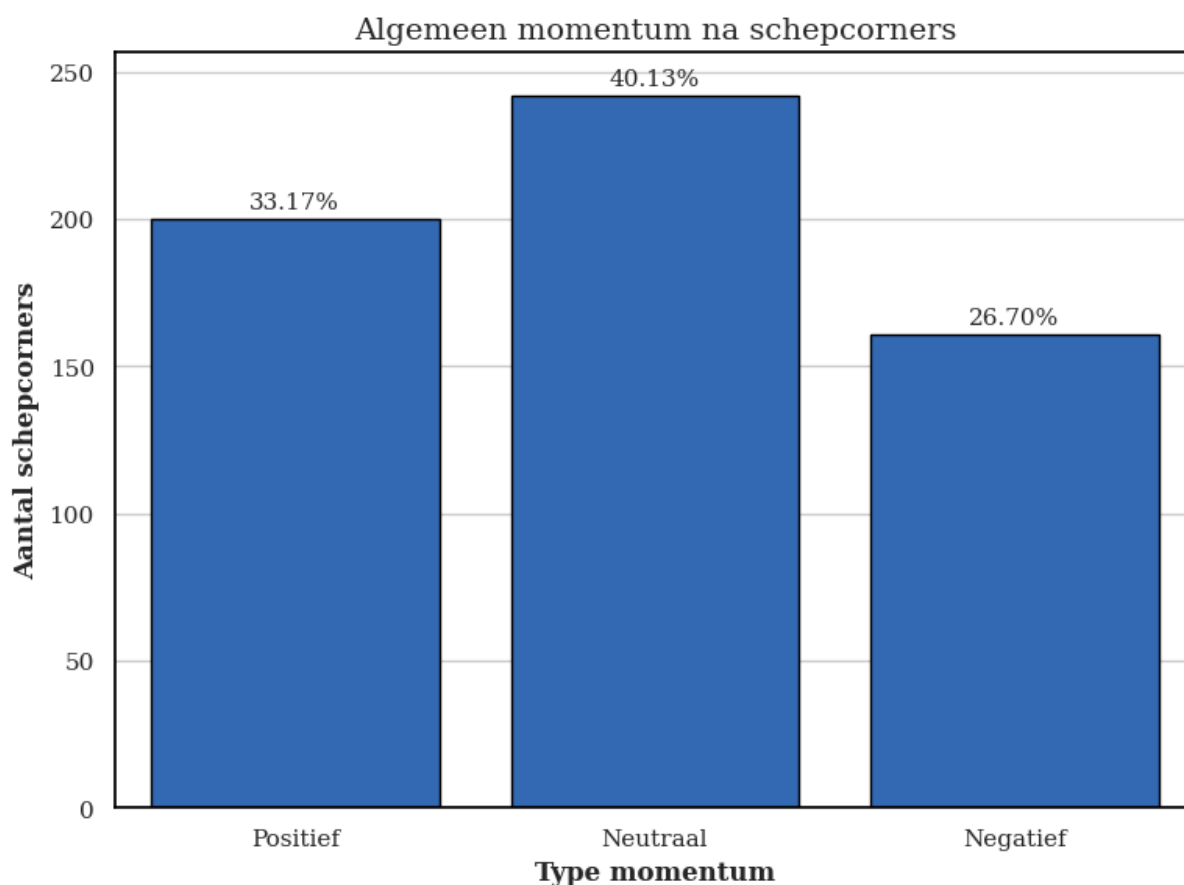
Tabel 8: Impact van schepcorners

De analyse toont aan dat in maar liefst 25,54% van de wedstrijden de uitkomst zou zijn gewijzigd als slechts één schepcorner een andere uitkomst mocht hebben. Met andere woorden: in één op de vier wedstrijden kan één enkele schepcorner het verschil maken. Dit onderstreept het directe, beslissende potentieel van deze spelhervatting. Wanneer men veronderstelt dat alle schepcorners een andere afloop kunnen hebben, stijgt dit aandeel tot 60,17%. Dit duidt op een aanzienlijke gecumuleerde impact van schepcorners binnen een wedstrijd. Ook in hypothetische extremen blijkt de impact niet te verwaarlozen. In 18,18% van de wedstrijden zou de eindstand anders geweest zijn als geen enkele schepcorner was omgezet in een doelpunt. Omgekeerd zou het scoren van alle schepcorners in 12,12% van de wedstrijden tot een andere uitkomst hebben geleid. Hoewel dit eerder theoretische scenario's zijn, benadrukken ze wel dat schepcorners aanzienlijk strategisch potentieel in zich dragen.

Samengevat suggereren deze resultaten dat schepcorners een niet te negeren invloed uitoefenen op het wedstrijdverloop. In een significant deel van de wedstrijden bepalen ze zelfs rechtstreeks het resultaat. Dit onderstreept het belang van het trainen en analyseren van deze spelhervatting binnen het minivoetbal.

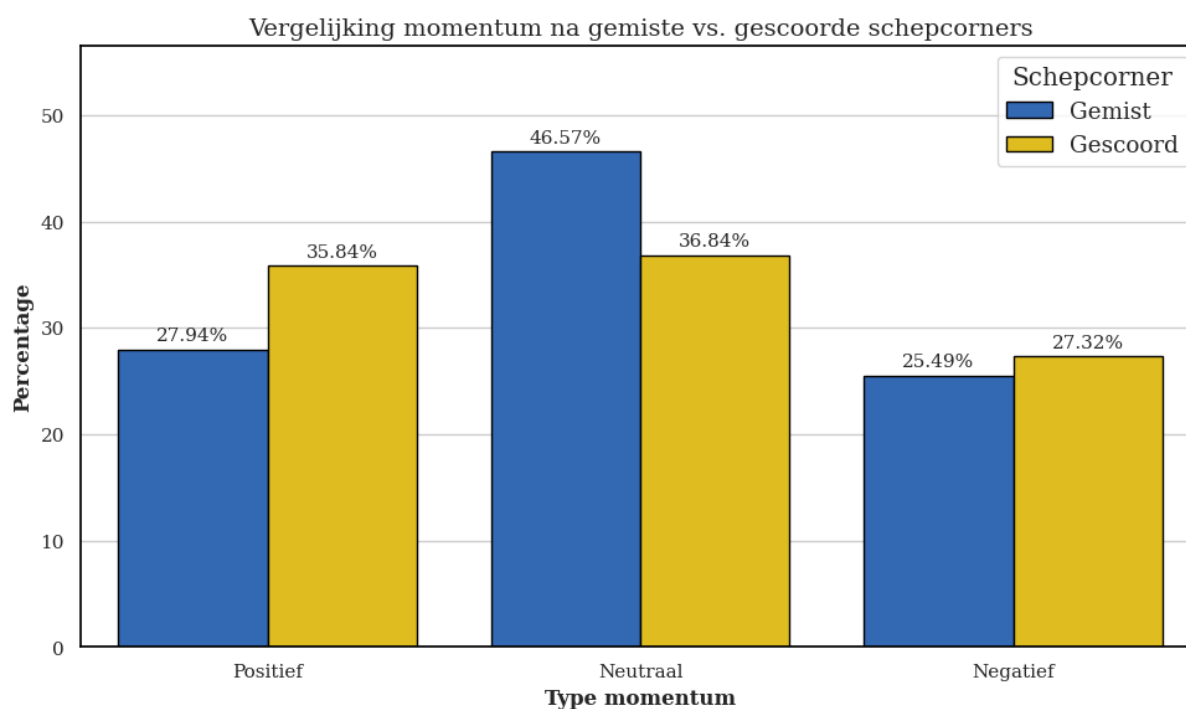
4.3 Momentum

Om te bepalen of het al dan niet scoren van schepcorners invloed heeft op het momentum van een wedstrijd, werd onderzocht welk team het meest scoorde in de fase onmiddellijk na de uitvoering van de schepcorner. De resultaten worden weergegeven in onderstaande figuur.



Figuur 2: Algemeen momentum na schepcorners

In 33,17% van de gevallen werd na een schepcorner een positief momentum waargenomen. In 40,13% bleef het momentum neutraal, terwijl in 26,70% sprake was van negatief momentum. Er werden dus meer positieve dan negatieve momentumverschuivingen geregistreerd na een schepcorner. Deze analyse betreft het algemene patroon na schepcorners, zonder onderscheid te maken tussen succesvolle en gemiste pogingen. In het hoofdstuk [Discussie](#) wordt besproken hoe deze bevinding mogelijk samenhangt met verschillen in het aantal schepcorners tussen winnende en verliezende teams. Om die algemene trend beter te begrijpen, werd vervolgens een onderscheid gemaakt tussen schepcorners die resulteerden in een doelpunt en diegene die gemist werden.



Figuur 3: Vergelijking momentum na gemiste vs. gescoorde schepcorners

In de bovenstaande grafiek wordt het verschil in momentum weergegeven na gescoorde en gemiste schepcorners. Voor beide groepen werd geanalyseerd welk type momentum (positief, neutraal of negatief) volgde op de spelhervatting. Wanneer de schepcorner werd omgezet in een doelpunt, trad in 35,84% van de gevallen positief momentum op. In 36,84% bleef het momentum neutraal en in 27,32% was er sprake van negatief momentum. Bij gemiste schepcorners werd in 27,94% van de gevallen een positief momentum waargenomen, in 46,57% een neutraal momentum en in 25,49% een negatief momentum.

De verdeling van momentum verschilt marginaal significant tussen gescoorde en gemiste schepcorners ($\chi^2 = 5,89$; $p = ,053$). Een gerichte toets op positief momentum toont een significant verschil: gescoorde schepcorners leiden vaker tot positief momentum dan gemiste schepcorners ($Z = 1,11$; $p = ,021$). Het scoren van schepcorners leidt dus tot een positief momentum voor de ploeg. Voor negatief momentum werd geen significant verschil vastgesteld ($Z = -0,48$; $p = ,684$). Op basis van deze resultaten kan niet worden aangetoond dat het missen van een schepcorner vaker tot negatief momentum leidt dan het scoren daarvan. Integendeel, in dit onderzoek werd zelfs minder negatief momentum waargenomen na gemiste dan na gescoorde schepcorners, wat tegen de intuïtie indruist.

Daarnaast werd het momentum vóór en na de uitvoering van een schepcorner met elkaar vergeleken. De resultaten zijn weergegeven in de onderstaande tabel.

	Verbetering	Ongewijzigd	Verslechtering
Score	34,34%	32,33%	33,33%
Missen	28,43%	38,24%	33,33%

Tabel 9: Vergelijking momentum voor en na schepcorner

Wanneer de schepcorner resulteerde in een doelpunt, werd in 34,34% van de gevallen een verbetering van het momentum vastgesteld. In 32,33% bleef het momentum ongewijzigd, en in 33,33% verslechterde het. Bij gemiste schepcorners lagen deze percentages respectievelijk op 28,43%, 38,24% en 33,33%.

De verschillen in de verdeling tussen beide groepen waren niet statistisch significant ($\chi^2 = 2,83$; $p = ,243$). Wel werd in een Z-test voor proportieverschillen een marginaal significant verschil vastgesteld voor de verbetering van het momentum na gescoorde schepcorners ($Z = 1,47$; $p = ,071$). Deze resultaten suggereren voorzichtig dat het momentum vaker verbetert na het scoren dan na het missen van een schepcorner, al vereist deze bevinding verdere bevestiging in toekomstig onderzoek.

Ten slotte werd onderzocht wat het effect was van het momentum voorafgaand aan het nemen van een schepcorner op de resulterende scorepercentages. Hieruit bleek dat positief, neutraal en negatief momentum respectievelijk voor een scorepercentage van 66,83%, 64,22% en 67,77% leidden. Opvallend is dat negatief momentum voorafgaand aan een schepcorner het hoogste scorepercentage opleverde, al liggen de percentages dicht bij elkaar en was het verschil tussen positief en negatief niet significant ($Z = 0,19$; $p = ,425$).

4.4 Impact regelwijziging

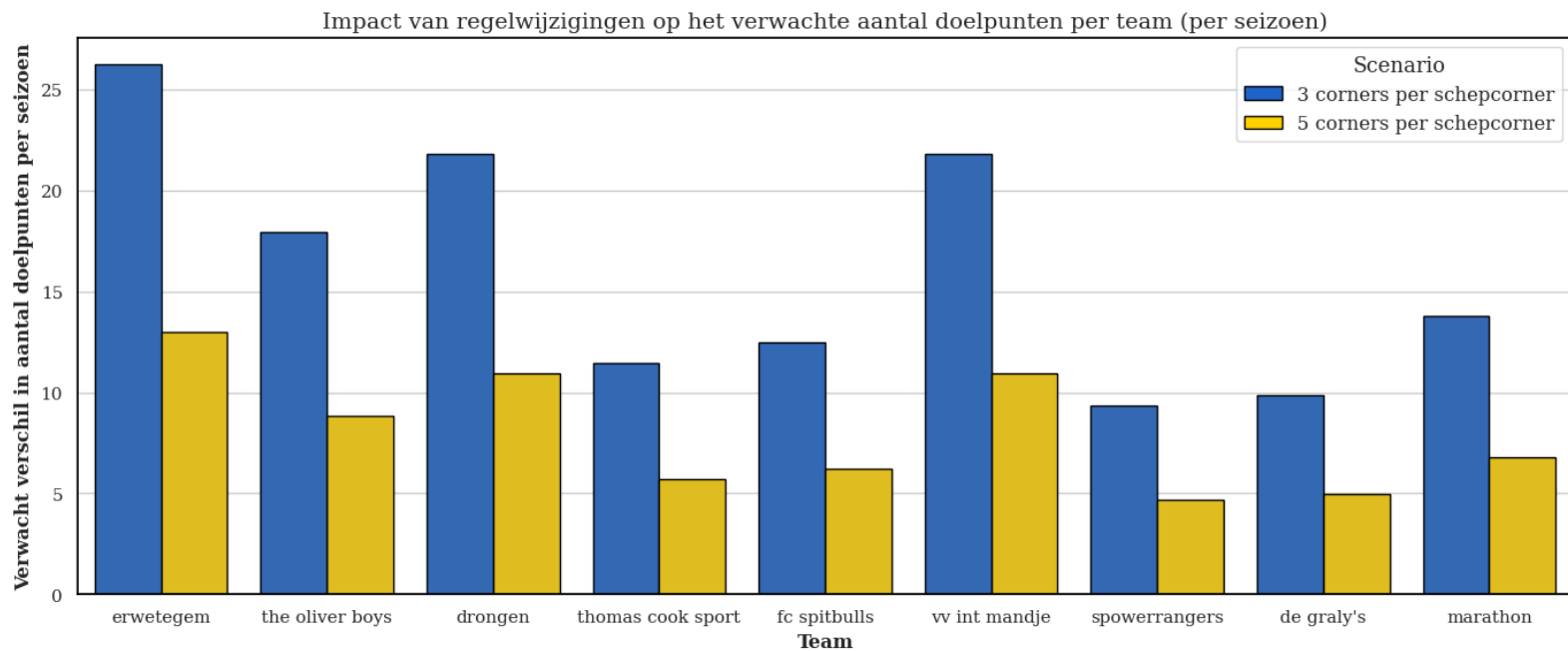
Voor deze analyse werd uitsluitend rekening gehouden met 396 van de 603 schepcorners. De overige 207 waarnemingen werden uitgesloten omdat er voor deze wedstrijden geen informatie beschikbaar was over het aantal cornerpunten bij het eindsignaal. Daardoor kon het totale aantal toegekende cornerpunten niet worden berekend, terwijl deze informatie noodzakelijk was om het effect van een regelwijziging te simuleren.

Regel (corners per schepcorner)	Aantal schepcorners	Verandering t.o.v. huidig
3	596	50.50% meer
4 (basis)	396	huidig
5	297	25% minder

Tabel 10: Impact van regelwijziging

Bovenstaande tabel toont het verwachte aantal schepcorners bij verschillende toekenningsregels. Momenteel wordt één schepcorner toegekend per vier hoekschoppen. De analyse geeft aan dat een wijziging naar drie hoekschoppen per schepcorner het aantal schepcorners met 50,5% zou verhogen, terwijl een wijziging naar vijf hoekschoppen per schepcorner tot een daling van 25% zou leiden.

Naast de globale analyse werd het effect van een regelwijziging ook onderzocht op teamniveau. De onderstaande figuur toont het verwachte verschil in het aantal doelpunten per seizoen voor alle teams waarvoor minstens tien schepcorners zijn opgenomen in de data. De visualisatie vergelijkt het aantal extra dan wel minder verwachte doelpunten per team onder de scenario's waarin een schepcorner wordt toegekend per drie of vijf corners.



Figuur 4: Extra of minder doelpunten op een seizoen per team bij regelwijziging

De relatieve verschillen tussen teams kunnen als volgt worden geïnterpreteerd. Erwetegem, dat het sterkst zou profiteren van een wijziging naar drie corners per schepcorner (of het meest zou inboeten bij vijf), combineert een hoog gemiddeld aantal corners per wedstrijd (2,20) met een uitzonderlijk hoge omzettingsgraad (90,91%). Teams die minder gevoelig zijn voor de wijziging, zoals Thomas Cook Sport of Marathon, onderscheiden zich door een lager gemiddeld aantal corners per wedstrijd (Thomas Cook Sport: 1,25; omzettingsgraad: 70,00%) of door een lagere omzettingsgraad bij een vergelijkbaar aantal corners (Marathon: 2,00; omzettingsgraad: 52,63%). Deze analyse toont aan dat de impact van een regelwijziging per team aanzienlijk kan variëren, afhankelijk van zowel het aantal verworven corners als de efficiëntie waarmee die worden benut.

Tot slot werd nagegaan welk effect een regelwijziging zou hebben op het behaalde aantal klassemmentpunten en de eindrangschikking van enkele ploegen. De onderstaande tabel toont het puntenverschil per ploeg onder de scenario's van respectievelijk drie en vijf corners per schepcorner, evenals de posities in het klassement.

Ploeg	Puntenverschil (3)	Puntenverschil (5)	Originele Ranking	Ranking (3)	Ranking (5)
Marathon	+4	0	2	1	2
FC Spitbulls	+7	+2	4	4	4
De Graly's	+2	-1	4	4	5
The Oliver Boys	-1	+1	8	9	8
Spowerrangers	+2	-2	5	4	5

Tabel 11: Puntenverschil en eindklassering bij regelwijziging

Uit de bovenstaande tabel blijkt dat een wijziging in de toekenningsregel niet alleen invloed heeft op het aantal doelpunten per seizoen, maar ook op de uiteindelijke rangschikking van sommige ploegen. Zo zou Marathon, dat in de oorspronkelijke situatie net naast de titel greep, bij een regel van drie corners per schepcorner als kampioen zijn geëindigd.

Het is logisch dat een regelwijziging niet voor elk team hetzelfde effect heeft. Een aanpassing naar drie corners per schepcorner leidt niet automatisch tot een stijging in het puntenaantal en een wijziging naar vijf corners resulteert niet per se in een daling. Deze contextafhankelijkheid wordt geïllustreerd door The Oliver Boys, die één punt verliezen in het scenario van drie corners per schepcorner. Dit verschil kan worden verklaard door een specifieke wedstrijd die oorspronkelijk op een gelijkspel eindigde, maar waarin de tegenstander onder de nieuwe regel een extra schepcorner verkrijgt en naar verwachting scoort. Dit zou hebben geleid tot verlies in plaats van een gelijkspel, en dus tot één punt minder in de eindstand.

FC Spitbulls daarentegen behaalt in het scenario van drie corners per schepcorner een duidelijk voordeel. Zij verkrijgen in dat geval vaker een extra schepcorner, wat resulteert in meer gewonnen wedstrijden en dus een stijging van het aantal behaalde punten. Wel dient opgemerkt te worden dat de toename in punten zich niet vertaalt in een hogere rangschikking.

Een belangrijke kanttekening bij deze analyse is dat de nieuwe rangschikkingen werden berekend zonder rekening te houden met het effect op de overige ploegen. Deze beperking was noodzakelijk, aangezien er voor veel andere teams onvoldoende wedstrijden beschikbaar waren. De ploegen die in deze analyse zijn opgenomen, beschikten elk over minstens zeven wedstrijden met volledige cornerinformatie, wat een redelijke basis vormt voor vergelijking.

Op basis van bovenstaande resultaten kan worden geconcludeerd dat een wijziging in de toekenningsregel voor schepcorners, zowel naar drie als naar vijf corners per schepcorner, leidt tot verschillende effecten per team. Sommige ploegen zouden punten winnen door de aangepaste regel, terwijl andere net punten verliezen. Hoewel het hier om een simulatie gaat met beperkingen in de dataset (zoals het ontbreken van gegevens voor alle ploegen), suggereren de resultaten wel degelijk dat het wijzigen van de huidige toekenningsregel een merkbare impact kan hebben op het eindklassement van het seizoen.

4.5 Belangrijke factoren achter het succes van de schepcorner

In dit onderdeel van de resultaten wordt onderzocht welke factoren het succes van een schepcorner mogelijk beïnvloeden. Eerst worden de resultaten per afzonderlijke variabele besproken, waarna wordt overgegaan tot de voorspellende modellen en de interpretatie van de variabelen die een belangrijke rol spelen in het verklaren van het succes. Hieronder volgt een algemene beschrijving van de dataset.

In totaal werden 231 wedstrijden geanalyseerd, waarin 603 schepcorners werden geregistreerd. Gemiddeld kwamen er 2,61 schepcorners per wedstrijd voor. Van deze pogingen werd 66,17% succesvol omgezet in een doelpunt. De onderstaande tabel toont de verdeling van de data per competitieniveau.

Niveau	wedstrijden	schepcorners	scorepercentage	Schepcorners per wedstrijd
Nationale 1	25	81	86.42%	3.24
Nationale 4	5	13	76.92%	2.6
Provinciale 2	1	4	<i>Te weinig data</i>	<i>Te weinig data</i>
Kern 1	16	42	69.05%	2.63
Kern 2	29	75	72.00%	2.59
Kern 3	75	182	63.74%	2.43
Kern 4	80	206	57.77%	2.58

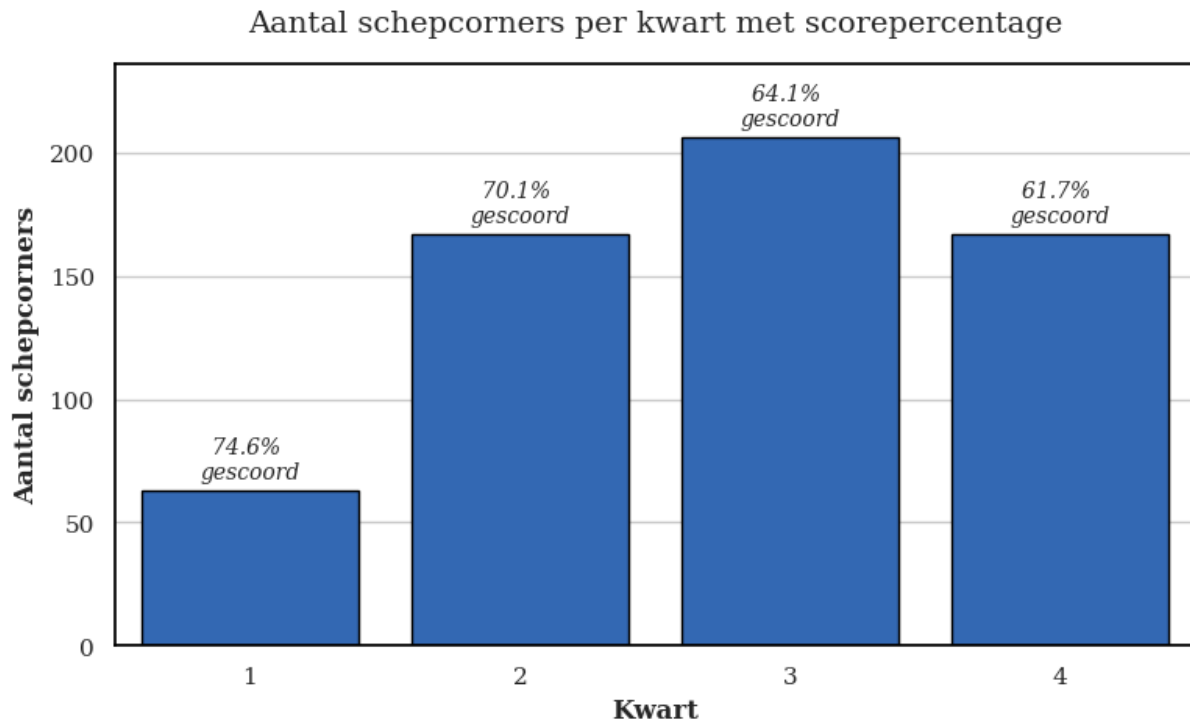
Tabel 12: Gegeven per competitieniveau

Uit de verdeling blijkt dat het grootste deel van de data afkomstig is uit de lagere kernreeksen (155 van de 231 wedstrijden). Dit weerspiegelt echter de structuur van de populatie, waarin zich meer ploegen bevinden in de lagere dan in de hogere reeksen. Zo zijn er bijvoorbeeld acht reeksen op kernniveau 4, tegenover slechts één op kernniveau 1. De categorie Provinciale 2 bevat slechts één wedstrijd, en daarom werden van dit niveau geen statistieken weergegeven.

Een eerste visuele indruk suggereert dat het scorepercentage stijgt naarmate het competitieniveau hoger is. Een chi-kwadraattoets bevestigde dat de verdeling niet uniform is ($\chi^2 = 26,81$; $p < ,001$). Om vervolgens na te gaan of deze stijging systematisch en lineair was, werd een Cochran-Armitage trendtest uitgevoerd. Deze test wees eveneens op een significante opwaartse trend in het scorepercentage naarmate het niveau stijgt ($Z = 4,41$; $p < ,001$).

Daarnaast werd onderzocht of het aantal schepcorners per wedstrijd verschilt tussen de niveaus. Een Kruskal-Wallis-toets toonde aan dat dit verschil significant is ($H = 15,87$; $p = ,014$). Aangezien het gemiddeld aantal schepcorners per wedstrijd in Nationale 1 opvallend hoger lag dan in de overige niveaus, werd aanvullend een gerichte t-toets uitgevoerd om na te gaan of dit verschil ook statistisch significant is. Hierbij werd Nationale 1 vergeleken met de andere niveaus tezamen. Deze analyse toonde aan dat het verschil significant is ($t = 3,89$; $p < ,001$), wat bevestigt dat er in Nationale 1 substantieel vaker gebruik wordt gemaakt van de schepcorner dan in de lagere reeksen. Dit verschil zou kunnen wijzen op een meer tactische benadering in Nationale 1, waarbij teams mogelijk vaker strategisch naar schepcorners toewerken.

4.5.1 Vermoeidheid



Figuur 5: Aantal schepcorners per kwart met scorepercentages

Een minivoetbalwedstrijd bestaat uit vier kwarten van telkens twaalf minuten. De bovenstaande figuur toont het aantal genomen schepcorners per kwart (y-as), met boven elke balk het overeenkomstige scorepercentage.

Ten eerste blijkt dat er in het eerste kwart aanzienlijk minder schepcorners worden genomen dan in de overige drie kwarten. Dit verschil werd getoetst met een Z-test voor proporties en bleek statistisch significant ($Z = -11,68$; $p < ,001$). Een plausibele verklaring hiervoor is dat bij aanvang van elke wedstrijd het aantal cornerpunten op nul staat, waardoor het in het eerste kwart minder waarschijnlijk is om reeds vier corners te verzamelen, de drempel die vereist is om een schepcorner te krijgen. In latere kwarten wordt deze drempel sneller bereikt, aangezien de cornerpunten van eerdere kwarten behouden blijven.

Ten tweede daalt het scorepercentage systematisch naarmate de wedstrijd vordert, van 74,6% in het eerste kwart naar 61,7% in het vierde. Om deze trend formeel te toetsen, werd een Cochran-Armitage trendtest uitgevoerd. Deze wees op een significante negatieve trend ($Z = -2,20$; $p = ,013$), wat suggereert dat de kans op succes bij een schepcorner afneemt naarmate de wedstrijd vordert. Een mogelijke verklaring hiervoor is het toenemende effect van vermoeidheid, wat de uitvoering en scherpheid van standaardsituaties kan beïnvloeden.

Bankbezetting	Aantal schepcorners	Aantal gescoord	Scorepercentage
Beide	2	2	100,00%
Kopper	65	41	63.08%
Schepper	87	51	67.93%
Niemand	449	305	58.62%

Tabel 13: Schepcornersucces verdeeld op bankbezetting

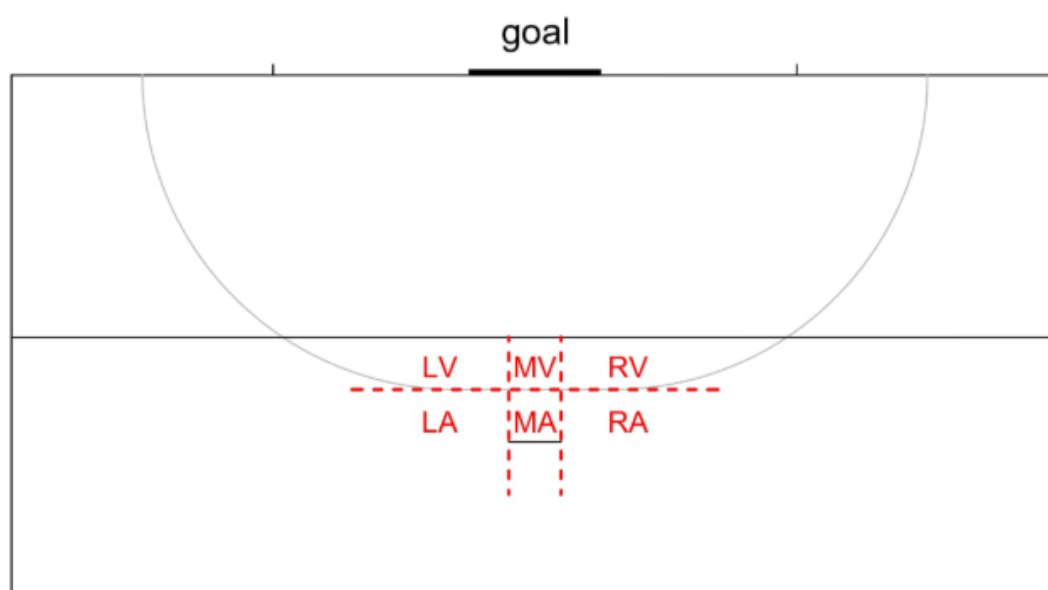
De bovenstaande tabel toont de verdeling van schepcorners naar gelang van de bankbezetting op het moment van uitvoering. In 74,5% van de gevallen werd de schepcorner genomen door spelers die zich reeds op het veld bevonden. Opvallend is dat het scorepercentage het laagst ligt bij die groep (58,6%), terwijl situaties waarbij een kopper of schepper van de bank kwam gemiddeld tot hogere conversie leidden. Deze bevinding sluit aan bij de hypothese dat vermoeidheid mogelijk een negatieve invloed heeft op de uitvoering van een schepcorner.

Een chi-kwadraattoets werd uitgevoerd om te onderzoeken of het scorepercentage tussen de vier bankopties verschilde van de normale verdeling. De test leverde echter geen significant resultaat op ($\chi^2 = 4,14$; $p = ,247$) waardoor niet kan worden geconcludeerd dat de uitkomst van een schepcorner afhangt van de bankconfiguratie.

4.5.2 Positionele kenmerken

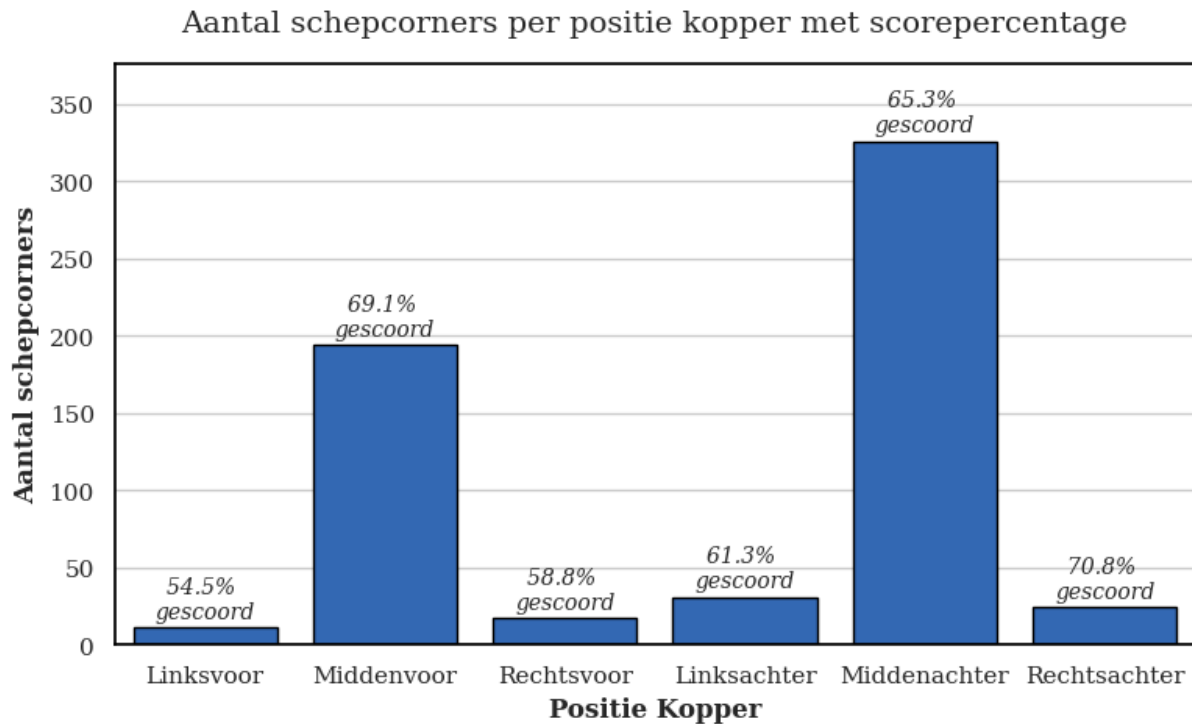
Verder blijkt dat het merendeel van de scheppers ervoor kiest om, vanuit het perspectief van de kopper, aan de rechterzijde van het doel te scheppen. Concreet gebeurde dit in 72,3% van de gevallen, tegenover 27,7% vanaf de linkerzijde. Het scorepercentage verschilt echter nauwelijks tussen beide zijden: 66,5% bij rechts geschepte schepcorners tegenover 65,3% bij links geschepte schepcorners.

Voor de positie van de kopper werden zes zones geïdentificeerd waarin deze zich kan opstellen. In de onderstaande figuur is te zien wat met elke zone wordt bedoeld. De afkortingen LV, MV en RV staan respectievelijk voor linksvoor, middenvoor en rechtsvoor. LA, MA en RA verwijzen naar linksachter, middenachter en rechtsachter.



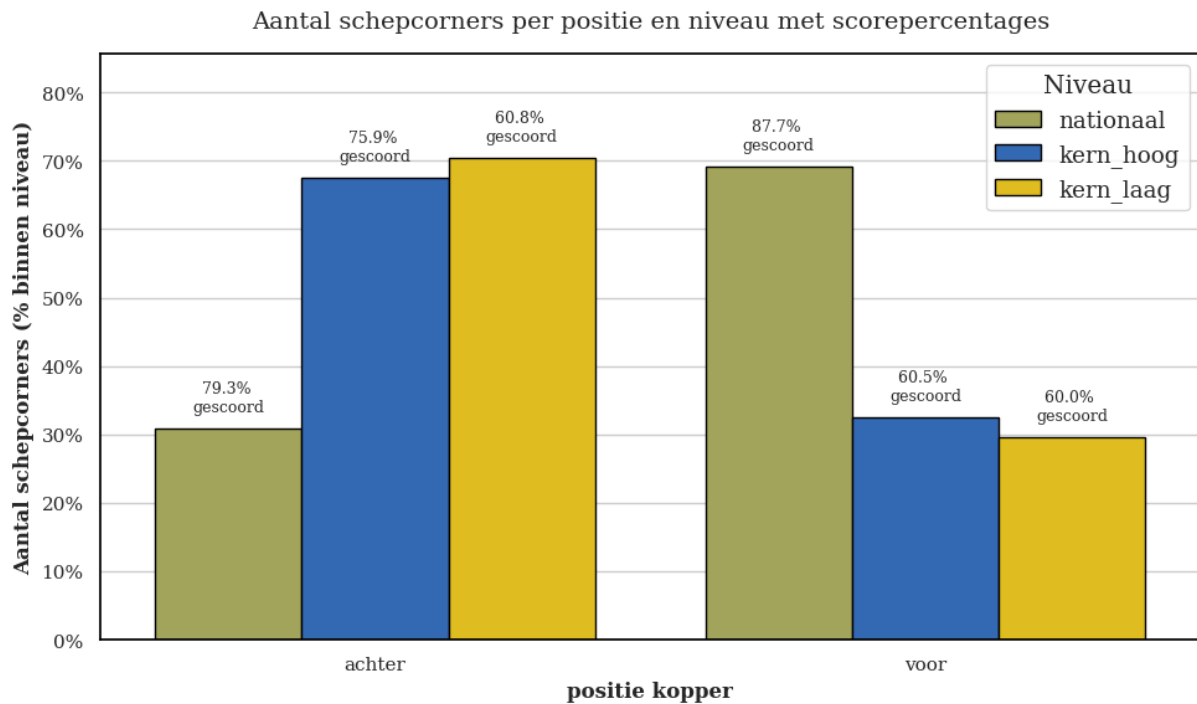
Figuur 6: Schets van mogelijke posities van de kopper

De onderstaande grafiek toont de verdeling van schepcorners over de verschillende posities die de koper kan innemen, met op de y-as het aantal schepcorners en boven elke balk het bijbehorende scorepercentage. Uit de figuur blijkt dat het merendeel van de kopers zich positioneert in het centrum, met name in de zone middenachter, gevolgd door middenvoor. Posities achteraan worden duidelijk frequenter gekozen dan posities vooraan.



Figuur 7: Aantal schepcorners per positie van de koper met de scorepercentages

Hoewel de scorepercentage enigszins variëren tussen de posities, van 54,5% (linksvoor) tot 70,8% (rechtsachter), zijn deze verschillen niet statistisch significant. Een chi-kwadraattoets voor onafhankelijkheid bevestigde dit ($\chi^2 = 2,47$; $p = ,756$), wat erop wijst dat de positie van de koper geen aantoonbaar effect heeft op het scorepercentage van de schepcorner.



Figuur 8: Scorepercentage en verdeling van posities over de verschillende competitieniveaus

De bovenstaande figuur toont per competitieniveau het aandeel schepcorners dat werd genomen vanuit een voorwaartse of achterwaartse positie van de koper, weergegeven op de y-as als percentage binnen elk niveau. De niveaus werden hiertoe herverdeeld in drie groepen: *kern_laag* (niveaus 3 en 4), *kern_hoog* (niveaus 1 en 2), en *nationaal* (alle niet-kernniveaus). De bijbehorende scorepercentages zijn weergegeven boven de balken.

Hoewel in de onderliggende data het merendeel van de koppers op elk niveau zich horizontaal in het midden positioneert (zie grafiek in de [Appendix](#)), valt vooral het verschil in verticale positionering per niveau op. Op nationaal niveau werd in 69,1% van de gevallen gekozen voor een voorwaartse positie, tegenover slechts 30,9% achteraan. In de kernreeksen is dit patroon duidelijk omgekeerd: in *kern_hoog* bevond de koper zich in 67,5% van de gevallen achteraan, en in *kern_laag* zelfs in 70,4%.

Deze verschillen in positievoorkeur blijken statistisch significant ($\chi^2 = 51,99$; $p < ,001$), wat suggereert dat spelers op verschillende competitieniveaus daadwerkelijk andere keuzes maken in hun positionering.

Wat betreft de scorepercentages zijn de verschillen tussen voor- en achterpositie binnen elk competitieniveau niet significant op het conventionele 5%-niveau, al blijven de onderliggende patronen opvallend. Op nationaal niveau is de effectiviteit van een voorwaartse positie duidelijk hoger (87,7%) dan die van een achterwaartse positie (79,3%), maar dit verschil is statistisch niet significant ($Z = 1,05$; $p = ,292$). In de kerncompetitie is het patroon omgekeerd: in *kern_hoog* bedraagt de scorekans 75,9% bij een achterwaartse positie tegenover 60,5% bij een voorwaartse positie. Dit verschil is marginaal significant ($Z = -1,72$; $p = ,085$). In *kern_laag* is de trend eveneens aanwezig, al liggen de percentages zeer dicht bij elkaar (60,8% vs. 60,0%), en is het verschil niet significant ($Z = -0,15$; $p = ,882$). Hoewel de verschillen in scoringskans niet statistisch significant zijn, blijft dit toch een opvallende observatie. Verdere analyse op basis van grotere steekproeven is aangewezen om deze tendensen betrouwbaarder te kunnen bevestigen.

4.5.3 Schepkwaliteit

	Aantal schepcorners	Scorepercentage
Verplaatsen naar achter		
Ja	61	9,8%
Nee	542	72,5%
Horizontaal verplaatsen		
Ja	136	17,6%
Nee	467	80,3%
Veel verplaatsen		
Ja	104	9,6%
Nee	499	78,0%
Goed geschept		
Ja	412	88,6%
Nee	191	17,8%

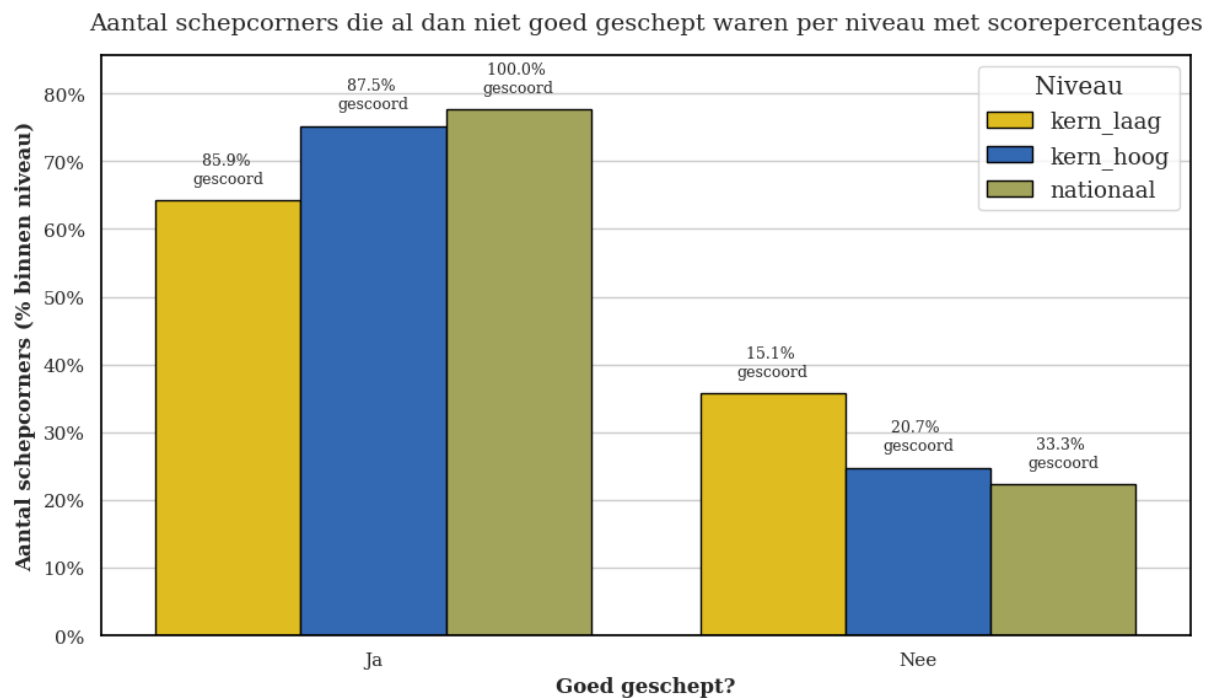
Tabel 14: Scorepercentage verdeeld op schepkwaliteit-indicatoren

De bovenstaande tabel toont de resultaten voor vier indicatoren die betrekking hebben op de uitvoeringskwaliteit van de schep. Hierbij werd nagegaan of de koper zich tijdens de actie moest verplaatsen naar achteren, opzij, of in sterke mate. De indicator *goed geschept* verwijst naar situaties waarin de koper zich niet in een van deze richtingen hoefde te verplaatsen en dus zonder, of met minimale voorwaartse verplaatsing, kon koppen.

De resultaten tonen aan dat de mate van verplaatsing van de koper sterk samenhangt met het scorepercentage. Wanneer er goed werd geschept, werd in 88,6% van de gevallen gescoord. Indien er wél verplaatsing vereist was, lag het scorepercentage aanzienlijk lager: bij verplaatsing naar achteren 9,8%, bij horizontale verplaatsing 17,6% en bij sterke verplaatsing 9,6%.

Deze verschillen zijn statistisch significant voor alle indicatoren. De chi-kwadraattoets leverde significante resultaten op voor *goed geschept* ($\chi^2(1) = 289.00$, $p < .001$), verplaatsing naar achteren ($\chi^2(1) = 93.43$, $p < .001$), horizontale verplaatsing ($\chi^2(1) = 181.90$, $p < .001$) en sterke verplaatsing ($\chi^2(1) = 176.52$, $p < .001$). Deze bevindingen onderstrepen dat de kwaliteit van de schep cruciaal is voor het succes van een schepcorner en bevestigen daarmee de visie van experts over wat een goed geschepte schepcorner kenmerkt.

Daarnaast werd onderzocht of er een verschil bestaat in de schepkwaliteit tussen de verschillende competitieniveaus.



Figuur 9: Schepkwaliteit vergeleken op competitieniveau

De bovenstaande grafiek toont de verdeling en effectiviteit van schepcorners in functie van de schepkwaliteit en het competitieniveau. Er wordt daarbij specifiek onderscheid gemaakt tussen schepcorners die als goed geschept of niet-goed geschept werden beoordeeld.

Beschrijvend valt op dat het aandeel schepcorners dat goed geschept werd systematisch toeneemt met het competitieniveau: van 64,2% bij *kern_laag* naar 75,2% bij *kern_hoog* tot 77,7% op *nationaal* niveau. Daarnaast blijkt dat ook het scorepercentage stijgt met het niveau, ongeacht of de schepcorner goed geschept werd of niet. Voor goed geschepte corners stijgt het scorepercentage van 85,9% (*kern_laag*) tot 100% (*nationaal*), terwijl ook bij niet-goed geschepte acties een duidelijke toename zichtbaar is: van 15,1% naar 33,3%.

Deze trends werden getoetst met de Cochran-Armitage trendtest. De analyse bevestigt dat het aandeel goed geschepte schepcorners significant toeneemt met het niveau ($Z = 2,95$; $p = ,002$). Ook het scorepercentage bij goed geschepte corners neemt significant toe ($Z = 3,08$; $p = ,001$), net als bij niet-goed geschepte corners ($Z = 2,03$; $p = ,021$).

Deze resultaten suggereren dat spelers op hogere niveaus niet alleen vaker correct scheppen, maar dat ook de effectiviteit van de kopper toeneemt, ongeacht de schepkwaliteit. Dit wijst op een hogere technische bekwaamheid van zowel scheppers als koppers naarmate het niveau stijgt.

4.5.4 Ervaring

Ervaringsniveau	1	2	3	4	5
Kopper	35.3%	29.6%	63.0%	74.0%	75.8%
Schepper	46.7%	23.8%	50.7%	75.8%	76.5%
Duo	46.9%	48.5%	53.9%	74.8%	79.0%
Algemeen	25.0%	40.5%	60.0%	72.7%	78.5%

Tabel 15: Scorepercentage verdeeld per ervaringsniveau

De bovenstaande tabel toont de trends in scorepercentage in functie van het ervaringsniveau. De categorie *algemeen* verwijst naar het gemiddelde van de overige ervaringsniveau variabelen. Dit verschilt van de variabele *duo*, die het gezamenlijke ervaringsniveau weergeeft van spelers die reeds samen een schepcorner uitvoerden. Een duo met twee individueel ervaren spelers kan dus alsnog een lage *duo*-score hebben indien zij weinig samen speelden.

Op beschrijvend niveau blijkt dat het scorepercentage in alle vier de categorieën toeneemt naarmate het ervaringsniveau stijgt. Zo stijgt de *algemeen*-categorie van 25,0% bij niveau 1 naar 78,5% bij niveau 5, met een gelijkaardig patroon voor het *duo* (46,9% naar 79,0%), de *kopper* (35,3% naar 75,8%) en de *schepper* (46,7% naar 76,5%).

Om de waargenomen stijgingen in scorepercentage formeel te toetsen, werd per categorie een Cochran-Armitage trendtest uitgevoerd. In alle vier de categorieën werd een significante positieve trend vastgesteld naarmate het ervaringsniveau toenam (alle p-waarden < .001). Deze uitkomsten bevestigen dat een hoger ervaringsniveau systematisch samenhangt met een grotere kans op succes bij een schepcorner, ongeacht of de ervaring individueel dan wel gedeeld is.

Deze bevindingen ondersteunen de stelling dat ervaring, zowel individueel als gezamenlijk, een sterke positieve invloed uitoefent op het succes van een schepcorner.

4.6 Voorspellende modellen voor schepcorner succes

In deze sectie worden de resultaten besproken van de classificatiemodellen die ontwikkeld zijn om het succes van een schepcorner te voorspellen. Zoals toegelicht in de methodologie werd gestart met een logistisch regressiemodel als interpreteerbare referentie. Vervolgens werden twee complexere modellen, Random Forest en XGBoost, toegevoegd om de voorspellende prestaties te verbeteren en de belangrijkste verklarende variabelen te identificeren via feature importance.

De analyse beoogt twee doelstellingen. Enerzijds wordt nagegaan welk model de beste voorspellende prestaties levert op ongeziene data, op basis van verschillende evaluatiecriteria. Anderzijds wordt onderzocht welke variabelen de grootste invloed uitoefenen op het al dan niet succesvol afronden van een schepcorner.

4.6.1 Voorspellende modellen met schepkwaliteit-indicatoren

	Accuracy	Precision	Recall	F1-score	AUC
Baseline LR	82,64%	79,00%	50%	88,27%	89,45%
Random Forest	83,47%	81,72%	59,52%	88,37%	91,83%
RF permutation importance	83,47%	81,05%	57,14%	88,51%	91,05%
RF Boruta	85,12%	83,52%	64,29%	89,41%	85,79%
XGBoost	81,82%	79,38%	52,38%	87,50%	89,63%
XGB permutation importance	84,30%	81,91%	59,52%	89,02%	90,42%
XGB Boruta	80,17%	77,78%	47,62%	86,52%	86,24%

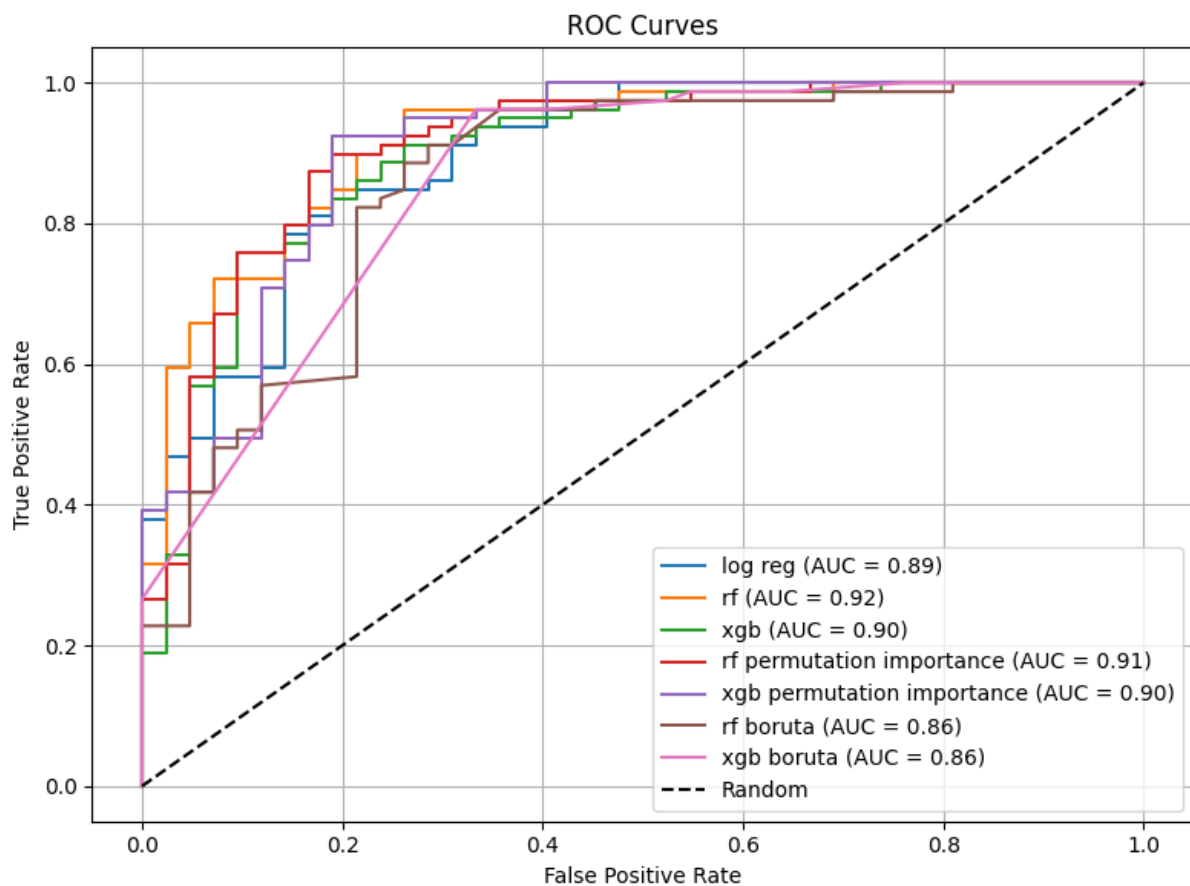
Tabel 16: Evaluatie van voorspellende modellen

De resultaten van de voorspellende modellen tonen aan dat het succes van een schepcorner goed voorspelbaar is op basis van de beschikbare variabelen. Alle modellen behalen een hoge AUC-score (tussen 0,85 en 0,92), wat wijst op een sterk onderscheidend vermogen tussen succesvolle en gemiste schepcorners over het volledige bereik van classificatiedrempels (Bradley, 1997).

Tegelijkertijd valt op dat de recall-waarden relatief laag blijven in alle modellen, met een maximum van 64,29% voor het Random Forest-model met Boruta-selectie. Dit toont aan dat het moeilijker is om gemiste schepcorners te voorspellen dan om gescoorde pogingen correct te identificeren. Deze bevinding is consistent met het probleem van class imbalance, waarbij modellen geneigd zijn te optimaliseren voor de dominante klasse (in dit geval: 'gescoord') ten koste van de minderheidsklasse (He & Garcia, 2009).

Wanneer de modellen onderling worden vergeleken, behaalt het standaard Random Forest-model zonder feature selection de hoogste AUC-score (91,83%), gevolgd door de variant met permutation importance (91,05%) en XGBoost met permutation importance (90,432). Een Delong-test toonde echter aan dat geen enkel model de baseline van het logistische regressiemodel significant over- of onderpresteerde. Het Random Forest-model met Boruta behaalt de hoogste scores op accuracy (85,12%), precision (83,52%), recall (64,29%) en F1-score (89,41%), maar heeft een lagere AUC-score (85,79%) in vergelijking met de andere modellen.

Een mogelijke verklaring voor de lagere AUC bij het Boruta-model is de beperkte informatiedichtheid: er werden slechts vijf variabelen geselecteerd, waaronder de drie schepkwaliteit-indicatoren, *algemene_ervaring* en *score_vershil_voor*. Door deze sterke reductie blijven enkel dominante signalen behouden, maar gaat nuance in de voorspellingen verloren. Hierdoor wordt het moeilijker om observaties correct te rangschikken over het volledige bereik aan drempelwaarden, wat resulteert in een lagere AUC-score. In de onderstaande ROC-grafiek is te zien dat de curves van het Boruta-model (zowel voor Random Forest als XGBoost) opvallend hoekiger verlopen.



Figuur 10: ROC curves van verschillende voorspellende modellen

Na evaluatie van de modellen werd het relatieve belang van de verklarende variabelen geïnterpreteerd op basis van het logistische regressiemodel. In Tabel 17 worden de variabelen weergegeven die minstens marginaal significant zijn op het 10%-niveau ($p < ,1$). De kolom 'Procentpunt verandering' geeft aan met hoeveel procentpunten de kans op een doelpunt verandert bij een toename van één eenheid in de betreffende variabele. Voor continue variabelen betreft dit een toename van één standaardafwijking, aangezien deze genormaliseerd werden met een *StandardScaler*. De standaardsituatie in het model is die waarin alle verklarende variabelen nul zijn (d.w.z. referentiecategorieën voor dummyvariabelen en gemiddelde waarden voor genormaliseerde continue variabelen).

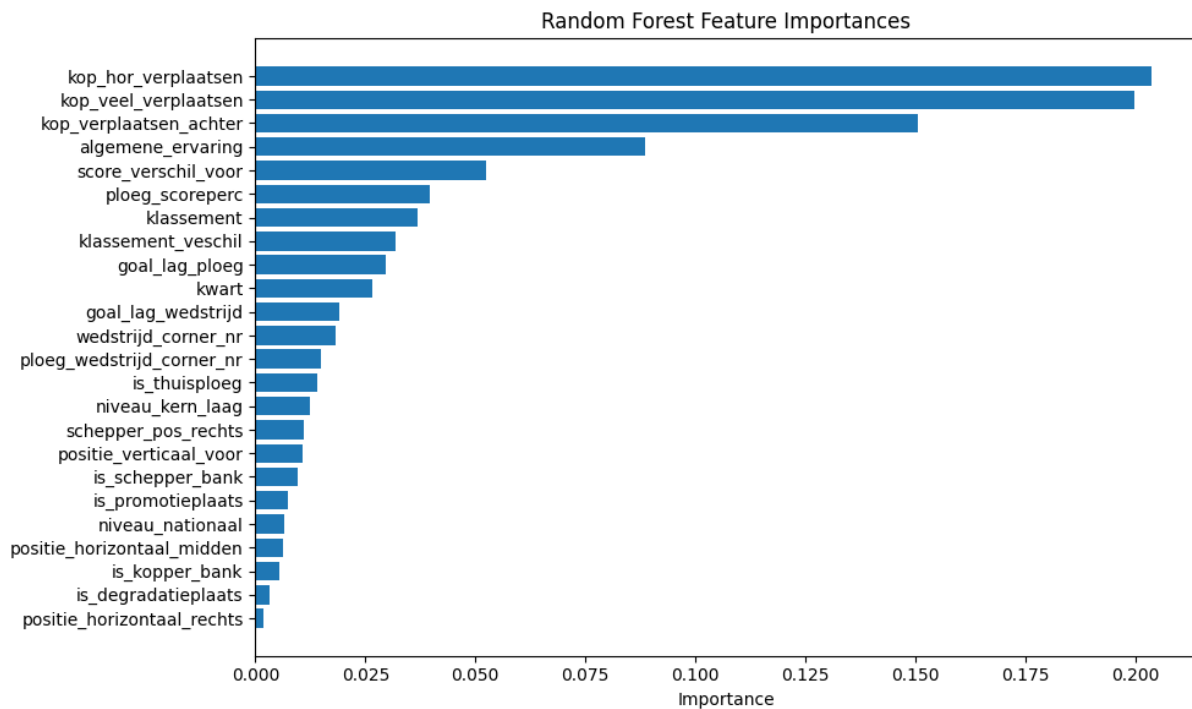
Variabele	Coëfficiënt	P-waarde	Procentpunt verandering
kop_verplaatsen_achter	-4,26	<,001	-87,26
kop_veel_verplaatsen	-2,98	<,001	-61,03
kop_hor_verplaatsen	-2,26	<,001	-46,35
goal_lag_ploeg	1,64	,021	33,54
score_verschil_voor	0,34	,049	6,94
positie_horizontaal_midden	-1,21	,055	-24,79
is_schepper_bank	-0,82	,062	-16,69
algemene_ervaring	0,35	,066	7,15
is_thuisploeg	-0,56	,093	-11,39

Tabel 17: Regressie-output

Wat onmiddellijk opvalt, is dat de drie schepkwaliteit-indicatoren een uitgesproken negatief effect hebben op de kans op een doelpunt. Dit effect is bovendien uitermate significant ($p < ,001$). Zowel achterwaartse verplaatsing, veelvuldige beweging als horizontale verplaatsing van de koper leiden tot een aanzienlijke daling in scoringskans, namelijk met respectievelijk met 87,26, 61,03 en 46,35 procentpunt. Deze resultaten bevestigen eerdere bevindingen en ondersteunen de hypothese dat een correct geschepte schepcorner leidt tot geen of minimale voorwaartse verplaatsing van de koper.

Daarnaast blijkt dat de kans op een doelpunt toeneemt naarmate het positieve scoreverschil groter is (+6,94 procentpunt), of wanneer het team over meer ervaring beschikt (+7,15 procentpunt). Het sterkste positieve effect wordt echter vastgesteld bij teams die bij de vorige schepcorner al hadden gescoord (+33,54 procentpunt), wat wijst op een mogelijk momentum- of vertrouwenseffect.

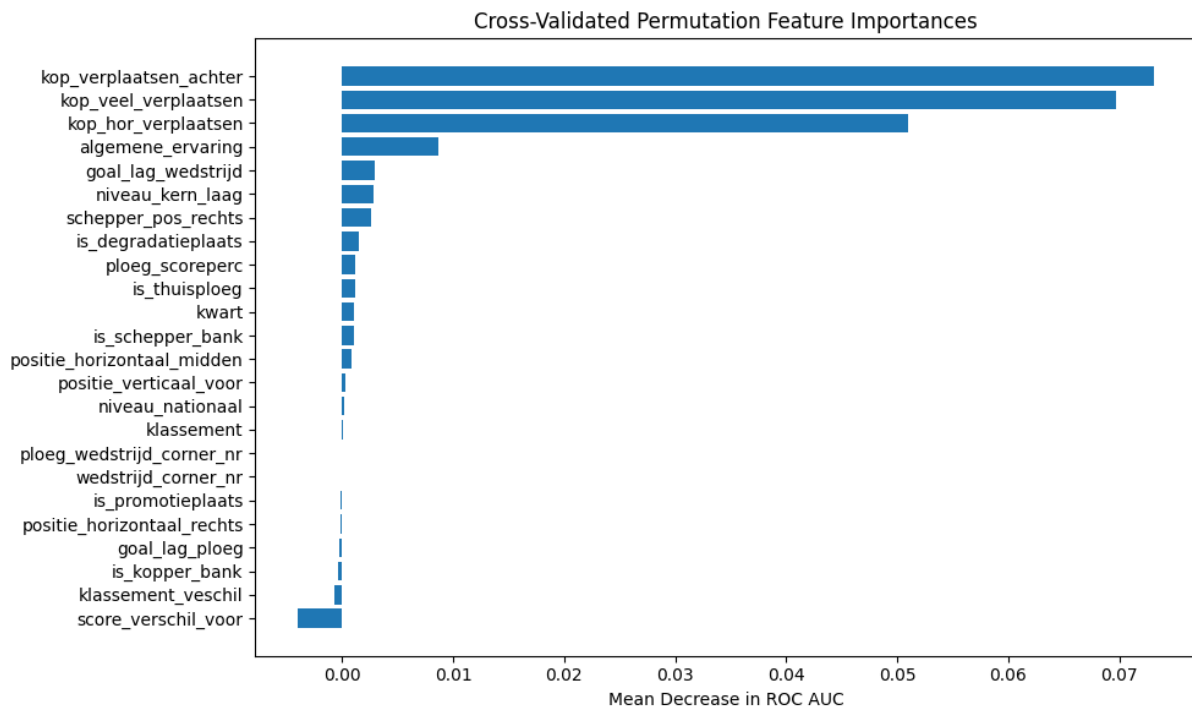
Enkele resultaten zijn minder intuïtief en bovendien slechts marginaal significant. Zo blijkt dat thuisspelende ploegen gemiddeld een lagere kans op een doelpunt hebben dan uitspelende ploegen (-11,39 procentpunt), wat onverwacht is. De meeste wedstrijden uit de dataset zijn namelijk afkomstig van kernploegen die geen vast thuisstadion hebben, waardoor het onderscheid tussen thuis- en uitwedstrijden vervaagt. Eveneens onverwacht is dat de aanwezigheid van de schepper op de bank leidt tot een lagere kans op een doelpunt (-16,69 procentpunt), terwijl eerdere beschrijvende analyses net een hogere scoringskans rapporteerden in situaties waarin de schepper van de bank kwam. Tot slot blijkt dat een positionering van de schepper in het midden van het veld minder effectief is dan een positie links (-24,79 procentpunt), wat niet strookt met eerdere descriptieve patronen.



Figuur 11: Feature importance van het Random Forest model

Aangezien het Random Forest-model de hoogste AUC-score behaalt, werd de bijbehorende feature importance-grafiek geraadpleegd om te bepalen welke variabelen het meest bijdragen aan de voorspellingen. Volgens dit model zijn de drie schepkwaliteit-indicatoren de meest informatieve variabelen voor het voorspellen van een doelpunt. Daarnaast blijken ook *algemene_ervaring* en *score_verschil_voor* belangrijke verklarende factoren binnen de modelstructuur. Deze variabelen dragen het meest bij aan het verminderen van de voorspellingsonzekerheid binnen de boomstructuur en worden dus frequent geselecteerd in beslissende splitsingen. Op basis hiervan kunnen we stellen dat het model deze variabelen beschouwt als bijzonder relevant voor het voorspellen van het al dan niet scoren uit een schepcorner.

Verder toont de Boruta-featureselectie aan dat uitsluitend dezelfde eerder besproken variabelen geselecteerd worden: de drie schepkwaliteit-indicatoren, *algemene_ervaring* en *score_verschil_voor*. Dit ondersteunt de hypothese dat deze variabelen modelmatig het meest informatief zijn voor het voorspellen van een doelpunt uit een schepcorner.



Figuur 12: Cross-Validated permutation feature importance

Wanneer gebruik wordt gemaakt van permutation importance als alternatieve methode voor featureselectie, blijkt uit bovenstaande figuur dat het willekeurig verstoren van de variabele *score_verschil_voor* paradoxaal genoeg leidt tot een lichte stijging van de gemiddelde AUC-score. Dit impliceert dat deze feature eerder ruis introduceert dan bijdraagt aan de voorspellingskracht van het model. Hoewel *score_verschil_voor* in eerdere analyses als significant naar voren kwam, suggereert dit resultaat dat de variabele mogelijk minder robuust of zelfs overbodig is binnen het Random Forest-model.

Ook de feature importance van het XGBoost-model en de bijbehorende selectiemethodes bevestigen dat vooral de schepkwaliteit-indicatoren een prominente rol spelen in het voorspellen van het al dan niet scoren uit een schepcorner. De bijbehorende visualisaties zijn terug te vinden in de [Appendix](#).

4.6.2 Voorspellende modellen zonder schepkwaliteit-indicatoren

Een belangrijk nadeel is echter dat de kwaliteit van de schep niet bekend is voordat de schepcorner genomen wordt. Daarom werden aanvullende modellen ontwikkeld zonder de schepkwaliteit-indicatoren. De prestaties van deze modellen vielen aanzienlijk lager uit, zoals blijkt uit Tabel 18.

	Accuracy	Precision	Recall	F1-score	AUC
Baseline LR	61,16%	66,33%	21,43%	73,45%	58,08%
Random Forest	65,29%	66,37%	9,52%	78,12%	58,02%
RF permutation importance	66,12%	67,98%	19,05%	77,84%	62,27%
RF Boruta	56,20%	63,00%	11,90%	70,39%	53,80%
XGBoost	64,46%	65,00%	0%	78,39%	58,11%
XGB permutation importance	65,29%	67,29%	16,67%	77,42%	57,96%
XGB Boruta	51,24%	77,78%	80,95%	48,70%	52,71%

Tabel 18: Evaluatie van voorspellende modellen zonder schepkwaliteit-indicatoren

Wanneer de schepkwaliteit-indicatoren worden weggelaten uit de modellen, blijkt dat de modelprestaties significant verslechteren ten opzichte van hun tegenhangers die deze informatie wel gebruiken. Het verschil in AUC werd getest via Delong en bleek voor alle modellen significant ($p < ,001$). Dit bevestigt dat schepkwaliteit een cruciale rol speelt bij het voorspellen van het al dan niet scoren uit een schepcorner.

Opvallend is dat het XGBoost-model zonder schepkwaliteitsinformatie een recall van 0% behaalt. Dit betekent dat het model geen enkele mislukte poging correct voorspelt, omdat het alle observaties als 'gescoord' classificeert. Deze overclassificatie van de positieve klasse leidt weliswaar tot een hoge F1-score (als gevolg van een hoge precision), maar ondermijnt de bruikbaarheid van het model in de praktijk.

Daartegenover staat het XGBoost-model met Boruta-featureselectie, dat een recall van 80,95% behaalt, maar dit resultaat bereikt door in de meeste gevallen een mislukte schepcorner te voorspellen. Dit wijst op een tegengesteld klassevooordeel. Dit wordt vermoedelijk verklaard door het feit dat Boruta in dit geval slechts één feature selecteerde (algemene_ervaring), wat resulteerde in een model met een zeer beperkte discriminatiecapaciteit.

Tot slot valt op dat het Random Forest model met permutation importance featureselectie het best presteert binnen deze groep. Dit model behaalt de hoogste AUC (62,27%) en accuracy (66,12%), wat suggereert dat deze selectieprocedure het meest robuust is wanneer schepkwaliteitinformatie ontbreekt. De gedetailleerde voorspellingen van elk model zijn terug te vinden in de [Appendix](#).

Binnen het logistisch regressiemodel zonder schepkwaliteit-indicatoren blijken meerdere variabelen een statistisch significante bijdrage te leveren aan de voorspelling van het al dan niet scoren uit een schepcorner. In onderstaande tabel worden de coëfficiënten, p-waarden en procentuele veranderingen in scoringskans weergegeven van de variabelen die minstens marginaal significant zijn.

Variabele	Coëfficiënt	P-waarde	Procentpunt verandering
algemene_ervaring	0,40	,001	8,51
kwart	-0,35	,015	-7,57
goal_lag_ploeg	0,87	,066	18,60
is_thuisploeg	-0,39	,074	-8,30
score_verschil_voor	0,20	,083	4,27

Tabel 19: Regressie-output zonder schepkwaliteit-indicatoren

De variabele *algemene_ervaring* toont een positief en sterk significant effect op de scoringskans ($p = ,001$): een toename van één standaardafwijking verhoogt de kans op een doelpunt met 8,51 procentpunt. Ook *score_verschil_voor* is positief geassocieerd met scoren (+4,27 procentpunt; $p = ,083$), hoewel het effect marginaal significant is. Verder blijkt uit het model dat wanneer een ploeg net voordien een doelpunt maakte (*goal_lag_ploeg*), dit samenhangt met een verhoogde kans op een volgend doelpunt (+18,60 procentpunt; $p = ,066$), mogelijk wijzend op een momentum-effect. Deze verbanden zijn consistent met de effecten die werden vastgesteld in het model met schepkwaliteit-indicatoren, al zijn de exacte effectgroottes licht verschillend. De stabiliteit van deze variabelen over beide modellen heen ondersteunt hun verklarende waarde, zelfs wanneer informatie over de uitvoering van de schepcorner ontbreekt.

Daarentegen blijkt dat het optreden als thuisploeg (*is_thuisploeg*) opnieuw geassocieerd is met een lagere kans op een doelpunt. Dit effect is marginaal significant (−8,30 procentpunt; $p = ,074$). Ook corners die plaatsvinden in latere kwarten gaan gepaard met een significante daling in scoringskans (−7,57 procentpunt, $p = ,015$), een patroon dat eerder ook in de descriptieve analyse werd vastgesteld.

Hoewel deze resultaten enkele interessante verbanden suggereren, is het belangrijk te benadrukken dat het model zonder schepkwaliteitsinformatie aanzienlijk slechter presteert. De verklaringskracht is beperkt en meerdere variabelen zijn slechts marginaal significant. De gerapporteerde effecten moeten dan ook met de nodige voorzichtigheid worden geïnterpreteerd.

Uit de analyse kunnen enkele conclusies worden getrokken over het belang van verklarende variabelen in deze context. Zonder schepkwaliteitsindicatoren vertrouwen zowel het Random Forest- als het XGBoost-model in sterke mate op *algemene_ervaring* als voorspeller van scoringskansen. Deze robuustheid wordt ondersteund door drie onafhankelijke analysemethoden: de standaard feature importance-scores, de permutation importance-analyse en de Boruta-featureselectie. In zowel de standaard- als de permutation-based feature importance-analyse werd *algemene_ervaring* consequent aangeduid als de meest bepalende variabele. Bovendien werd deze variabele in beide modellen als enige geselecteerd door het Boruta-algoritme. Hoewel andere variabelen, zoals *score_verschil_voor* en *niveau_nationaal*, eveneens hoog scoren in de importance-ranking, blijken ze minder bepalend te zijn voor de modelstructuur. De bijbehorende grafieken van de feature importance- en permutation importance-analyses zijn opgenomen in de Appendix.

5 Discussie

Dit onderzoek had twee centrale doelstellingen. Enerzijds werd beoogd bij te dragen aan de methodologische aanpak van onderzoek naar nichesporten, met bijzondere aandacht voor de praktische uitdagingen bij dataverzameling. Anderzijds was het de bedoeling een omvattend inzicht te verwerven in de rol van de schepcorner binnen het minivoetbal. In dat kader werden vier onderzoeksvragen geformuleerd: (1) wat is de impact van de schepcorner op het wedstrijdverloop, (2) in welke mate creëert de schepcorner momentum, (3) welk effect zou een wijziging in de toekenningsregel hebben, en (4) wat zijn de bepalende factoren voor een succesvolle uitvoering?

Uit de dataverzameling bleek dat online enquêtes in deze context, ondanks hun praktische voordelen zoals lage kosten en brede verspreidbaarheid, geconfronteerd werden met een zeer lage voltooiingsgraad. Slechts 28% van de externe respondenten die de enquête openden, vulde deze volledig en correct in. Voor veldonderzoek naar nichesporten lijkt het daarom aangewezen om, ondanks de hogere praktische drempel, het gebruik van papieren enquêtes te overwegen.

De impactanalyse toonde aan dat schepcorners een centrale rol spelen in het verloop van een minivoetbalwedstrijd. Slechts één van de 231 geanalyseerde wedstrijden bevatte geen enkele schepcorner. Daarbovenop kunnen teams er in de praktijk van uitgaan dat zij in negen van de tien wedstrijden minstens één schepcorner zullen uitvoeren. In 60,17% van de wedstrijden zou de uitslag anders zijn geweest (bijvoorbeeld een andere winnaar of een gelijkspel) indien de uitkomsten van schepcorners anders waren verlopen.

De momentumanalyse wees uit dat het verkrijgen van een schepcorner vaker tot positief dan tot negatief momentum leidt, ongeacht of er gescoord wordt. Wanneer effectief werd gescoord, leidde dit significant vaker tot positief momentum dan wanneer de kans werd gemist. Daarboven was het verschil van verbetering van momentum tussen scoren en missen marginaal significant. Gemiste kansen leidden daarentegen niet significant vaker tot negatief momentum of een verslechtering van de spelsituatie.

Daarnaast werd nagegaan wat het effect zou zijn van een wijziging in de toekenningsregel van de schepcorner. Indien de drempel wordt aangepast van één schepcorner per vier corners naar drie, zou het totale aantal schepcorners toenemen met 50,5%. Een verhoging van de drempel naar vijf corners zou daarentegen leiden tot een daling van 25%.

Uit de analyse van verklarende variabelen bleek dat vooral de schepkwaliteit-indicatoren sterk samenhangen met het al dan niet succesvol afronden van een schepcorner. Wanneer een schepcorner als *goed geschept* werd geclassificeerd (d.w.z. de kopper hoefde zich niet naar achteren, opzij of veel te verplaatsen), werd in 88,6% van de gevallen een doelpunt gescoord. Ook het ervaringsniveau van de spelers en het competitieniveau bleken significant samen te hangen met het scorepercentage: hoe meer ervaring en hoe hoger het niveau, hoe groter de kans op succes. Daarnaast nam het scorepercentage af naarmate de wedstrijd vorderde (van 74,6% in het eerste kwart naar 61,7% in het vierde kwart). Tot slot bleek dat teams die op het moment van de schepcorner voor stonden evenals teams die de vorige schepcorner hadden gescoord, een hogere scoringskans hadden.

5.1 Interpretatie van de resultaten

5.1.1 Dataverzameling en enquêtes

Zoals eerder vermeld werd de online enquête verstuurd naar alle minivoetbalploegen in de kerncompetitie van Gent. De respons op deze digitale versie bleef echter beperkt: slechts een klein aantal spelers opende de enquête en van deze groep werd het merendeel (72%) van de formulieren onvolledig of niet ingevuld. Naar aanleiding hiervan werden enkele spelers persoonlijk bevraagd om de oorzaken te achterhalen. Uit deze gesprekken kwamen meerdere verklaringen naar voren. De meest voorkomende reden was dat spelers de enquête eenvoudigweg vergaten, ondanks dat ze op de hoogte waren van het onderzoek. Aangezien minivoetbal op het kernniveau voornamelijk recreatief wordt gespeeld, bleek het invullen van een digitale vragenlijst geen prioriteit. Andere praktische obstakels waren het ontbreken van mobiel bereik in sporthallen, het feit dat spelers hun telefoon in de kleedkamer lieten, en de moeilijkheid om tijdens de wedstrijd actief een enquête op één smartphone te beheren, aangezien de eigenaar van de smartphone natuurlijk niet tijdens elke schepcorner op de bank zit.

Deze bevindingen bevestigen eerdere literatuur over lage responsgraden bij beperkte toegang tot technologie (Fricker & Schonlau, 2002; Ball, 2019), wat in dit onderzoek blijkt uit het vergeten of achterlaten van de smartphone. Daarnaast toont dit onderzoek aan dat ook locatiegebonden factoren, zoals slecht bereik in sporthallen, de respons verder beperken.

Om deze praktische beperkingen te ondervangen, werd besloten om over te schakelen op een papieren versie van de enquête. Deze aanpak bleek efficiënter: het fysieke formulier fungeerde als visuele herinnering, er was geen internetverbinding of toestel vereist, en dankzij de aanwezigheid van pennen moesten de spelers zelf niets meenemen om de enquête in te vullen. Bovendien kon de vragenlijst door eender wie ingevuld worden, waardoor een wisselende bank niet voor moeilijkheden zorgt.

Twee versies van het papieren formulier werden verspreid, met een klein verschil in de volgorde van de vragen. In de eerste, chronologisch opgebouwde versie werd informatie over de eindsituatie (zoals de eindscore en het aantal cornerpunten) pas op de achterzijde bevraagd, terwijl deze in de tweede versie al op de voorzijde van het formulier was opgenomen. Deze ogenschijnlijk kleine wijziging had een merkbaar effect op de datakwaliteit: in de tweede versie werden aanzienlijk minder gegevens over de eindsituatie overgeslagen. Dit sluit aan bij de bevindingen van Fanning (2005) die stelt dat zelfs beperkte aanpassingen in de opbouw van een vragenlijst, zoals de positionering van items, een grote impact kunnen hebben op de volledigheid en nauwkeurigheid van de antwoorden. Uit gesprekken met respondenten bleek dat dit verschil te verklaren was doordat in de meeste wedstrijden minder dan zes schepcorners voorkomen, waardoor de achterzijde van het formulier vaak niet werd opgemerkt. Door de eindvragen op de voorzijde te plaatsen, werden respondenten eerder herinnerd aan het invullen ervan. Wanneer dit alsnog werd vergeten, bood het omdraaien van het blad bij het inleveren van de enquête een visuele herinnering.

5.1.2 Impact

De resultaten van dit onderzoek wijzen op een aanzienlijke invloed van de schepcorner op het wedstrijdverloop in het minivoetbal. Uit de impactanalyse blijkt dat in 25,54% van de wedstrijden de einduitslag zou veranderen indien de uitkomst van slechts één schepcorner zou worden gewijzigd. Wanneer alle schepcorners binnen een wedstrijd van resultaat kunnen veranderen, loopt dit aandeel zelfs op tot 60,17%. Deze bevindingen onderstrepen het belang van de schepcorner als spelsituatie en suggereren dat ze een cruciale rol speelt in het bepalen van de uiteindelijke winnaar.

5.1.3 Momentum

Wat betreft het effect op momentum blijkt dat het verkrijgen van een schepcorner vaker leidt tot positief dan tot negatief momentum, ongeacht of de kans effectief werd omgezet in een doelpunt. Een verklaring hiervoor is dat winnende teams gemiddeld meer schepcorners krijgen dan teams die gelijkspelen of verliezen. Concreet krijgen winnende, gelijkspelende en verliezende ploegen gemiddeld respectievelijk 1,48, 1,23 en 1,15 schepcorners per wedstrijd. Aangezien winnende teams per definitie meer doelpunten scoren dan hun tegenstanders, is het plausibel dat het momentum na een schepcorner vaker in hun voordeel uitvalt, los van de directe uitkomst van de spelsituatie. Daarnaast blijkt dat het scoren van een schepcorner vaker leidt tot positief momentum, of minstens tot een verbetering van het momentum, in vergelijking met het missen ervan. Een geslaagde schepcorner levert dus niet alleen een doelpunt op, maar lijkt het team ook aan te zetten tot offensief overwicht in de daaropvolgende fase. In combinatie met de resultaten uit de impactanalyse suggereert dit dat een succesvolle schepcorner een doorslaggevende rol kan spelen in het verloop van een wedstrijd, zeker in evenwichtige duels.

5.1.4 Bepalende factoren voor schepcornersucces

Zoals eerder aangetoond in de resultaten, blijkt de kwaliteit van de schep de belangrijkste determinant voor het succes van een schepcorner. Met name de indicatoren die aangeven of de koper zich veel, horizontaal of naar achteren moet verplaatsen, vertonen een sterke negatieve samenhang met de scoringskans. Wanneer geen van deze verplaatsingen nodig is, wordt in 88,6% van de gevallen een doelpunt gescoord. Dit bevestigt ook het oordeel van experts, die stellen dat een schepcorner als kwalitatief goed wordt beschouwd wanneer de koper minimaal voorwaarts of niet hoeft te bewegen. Daarnaast speelt ook het competitieniveau een significante rol: hoe hoger het niveau, hoe groter de kans op een succesvolle afronding. Interessant is dat deze twee factoren elkaar versterken. Zo blijkt dat de schepkwaliteit verbetert naarmate het niveau stijgt, wat wijst op een grotere technische vaardigheid van de schepper. Bovendien benutten koppers op hoger niveau schepcorners vaker, ongeacht de schepkwaliteit, wat wijst op een hoger individueel afwerkingsvermogen.

Ook de positie van de koper verschilt systematisch tussen niveaus. Op nationaal niveau wordt vaker gekozen voor een voorwaartse positie dicht bij het doel, terwijl spelers op kernniveau zich doorgaans meer naar achteren opstellen. Deze positionele verschillen gaan gepaard met variatie in scoringskansen. Op nationaal niveau blijken voorwaartse posities succesvoller, terwijl op kernniveau de achterwaartse posities vaker leiden tot een doelpunt. Een plausibele verklaring hiervoor is de afweging tussen foutenmarge en technische precisie van de schepper. Wanneer de koper zich dicht bij de rand van het strafschoopgebied positioneert, dus vooraan, wordt de kans op scoren enerzijds groter aangezien de afstand tussen de koper en het doel kleiner is. Daar tegenover staat dat de foutmarge voor de schep kleiner is. Wordt de bal te kort of te dicht bij het doel geschept, dan moet de koper het strafschoopgebied betreden om de bal te kunnen koppelen, wat leidt tot een ongeldige schepcorner. Anderzijds kan een te ver geschepte bal ertoe leiden dat de koper zich naar achteren moet verplaatsen, wat, zoals aangetoond, sterk samenhangt met een lage omzettingsgraad. Hierdoor moet de schepper de bal met grote precisie naar de exacte positie van de koper spelen. Wanneer de koper zich echter meer naar achter positioneert, is er een grotere buffer tussen de koper en het strafschoopgebied. De schepper kan zich dan richten op een punt iets vóór de positie van de koper. Hierdoor kan de foutmarge van de schep toenemen zonder dat de koper zich naar achteren moet verplaatsen of het strafschoopgebied moet betreden.

De hogere schepkwaliteit op nationaal niveau toont aan dat technisch vaardige spelers in staat zijn om met voldoende precisie te scheppen naar een voorwaarts gepositioneerde koper. Hierdoor kunnen zij het risico op een te korte of te verre schep aanzienlijk beperken. Op lager niveau ligt de schepkwaliteit merkbaar lager, wat gepaard gaat met een grotere onnauwkeurigheid. Spelers kiezen er daarom vaker voor om de koper verder van het doel te positioneren, zodat er meer foutmarge ontstaat bij de uitvoering. Dit verklaart waarom een achterwaartse positie op kernniveau relatief vaker succesvol is, terwijl op nationaal niveau net de voorwaartse positionering tot betere resultaten leidt.

5.2 Praktische implicaties van de onderzoeksresultaten

Dit onderzoek richtte zich enerzijds op het verzamelen van data binnen een nichesporten, en anderzijds op de inhoudelijke analyse van de schepcorner in het minivoetbal. Hieronder worden de voornaamste praktische implicaties van de resultaten uit deze onderzoeken besproken.

5.2.1 Dataverzameling in de nichesport

In tegenstelling tot onze initiële verwachtingen, die ondersteunt werden door literatuur (Siva Durga Prasad Nayak & Narayan, 2019), bleken online enquêtes niet bij uitstek geschikt voor veldonderzoek in het minivoetbal. Aangezien minivoetbal, net als veel andere nichesporten, doorgaans in sporthallen wordt beoefend, lijkt het aannemelijk dat deze bevinding overdraagbaar is naar andere gelijkaardige contexten. Bij het opzetten van dataverzameling in nichesporten dienen enkele praktische factoren in overweging genomen te worden: beschikken respondenten over een geschikt apparaat om de enquête in te vullen, is er voldoende netwerkbereik, en kan de enquête effectief door één persoon worden ingevuld?

Daarnaast is het belangrijk te erkennen dat er in veel nichesporten geen doorgedreven professionalisering aanwezig is. De beoefening is doorgaans recreatief, wat betekent dat onderzoekers sterk afhankelijk zijn van de bereidwilligheid van spelers. Om de voltooiingsgraad en datakwaliteit te maximaliseren is het essentieel om het invullen van de enquête zo laagdrempelig mogelijk te maken. Een papieren vragenlijst dient dan ook niet bij voorbaat uitgesloten te worden. Het is aangewezen een afweging te maken tussen de lage kosten, brede verspreidbaarheid en het verwerkingsgemak van een online enquête, en de vaak hogere datakwaliteit, voltooiingsgraad en gebruiksvriendelijkheid van de papieren variant.

Bij gebruik van een papieren vragenlijst blijkt bovendien de volgorde van de vragen van belang. Het is aangewezen om te starten met algemene vragen en pas daarna over te gaan op het repetitieve gedeelte (bijvoorbeeld het invullen van individuele spelsituaties). Zelfs wanneer dit afwijkt van het chronologische verloop van de wedstrijd en dus de manier waarop de respondent de vragen kan invullen. Op die manier wordt gegarandeerd dat de respondent minstens alle vragen te zien krijgt, wat de volledigheid en kwaliteit van de gegevens ten goede komt.

5.2.2 De schepcorner in het minivoetbal

Zoals uit de resultaten blijkt, heeft de schepcorner een aanzienlijke impact op de wedstrijduitslag en leidt het scoren ervan tot proportioneel meer positief momentum. Deze stilstaande fase verdient dan ook meer aandacht binnen de voorbereiding van minivoetbal wedstrijden. Op basis van de bevindingen van dit onderzoek kunnen enkele gerichte aanbevelingen worden geformuleerd.

Ten eerste blijkt ervaring een cruciale rol te spelen. Teams wordt daarom aangeraden om zoveel mogelijk met vaste duo's te werken bij het nemen van schepcorners. Voor de positie van de schepper lijkt het minder relevant vanaf welke zijde de schepcorner wordt genomen; belangrijker is het om consistent te blijven in de keuze en uitvoering. Voor de positionering van de koper kunnen echter wel aanbevelingen worden gedaan. Beginnende teams en teams op een lager niveau doen er goed aan om een ruime afstand tot het strafschoopgebied te bewaren en zich meer naar achteren te positioneren. Dit vergroot de foutenmarge voor de schepper en maakt het voor de koper makkelijker om zich aan te passen aan een iets minder precieze schep. Naarmate de schepper meer ervaring en consistentie ontwikkelt, wordt een positie dicht bij het doel aanbevolen, aangezien de scoringskans toeneemt wanneer de koper vanop kortere afstand kan afwerken.

Tot slot biedt dit onderzoek ook handvatten voor beleidsmatige reflectie. Voor Voetbal Vlaanderen, de bevoegde organisator van het minivoetbal in Vlaanderen, blijkt dat kleine aanpassingen in de spelregels rond schepcorners een aanzienlijk effect kunnen hebben op het spelverloop. Indien één schepcorner wordt toegekend per drie corners, stijgt het aantal schepcorners met 50,50%. Een verhoging naar vijf corners per schepcorner resulteert daarentegen in een daling van 25%. Bovendien tonen simulaties aan dat dergelijke aanpassingen niet alleen het spelbeeld beïnvloeden, maar ook een merkbare impact hebben op wedstrijdresultaten en eindklassementen. In meerdere gevallen leidde een wijziging van de toekenningsregel tot een andere positie in het klassement. Dergelijke relatief eenvoudige ingrepen bieden dus een krachtig beleidsinstrument om de rol van de schepcorner te versterken of juist te beperken, afhankelijk van de nagestreefde doelstellingen van de regelgever.

6 Beperkingen en toekomstig onderzoek

Dit onderzoek vormt een eerste verkennende studie naar dataverzameling en analyse binnen een nichesport. Door het gebrek aan gelijkaardige voorgaande studies moeten de voorgestelde methoden en inzichten nog worden gevalideerd in vergelijkbare sportomgevingen. Het onderzoek biedt dan ook een vertrekpunt voor verdere academische reflectie en empirische uitbreiding binnen het veld van analytisch onderzoek naar nichesporten.

Een eerste methodologische beperking betreft de steekproef. Er werd gebruikgemaakt van een gemakkelijkssteekproef met als doel een maximaal aantal waarnemingen te verzamelen. Dit leidde echter tot een dataset waarin vooral ploegen uit de lagere kernniveaus oververtegenwoordigd zijn, met enkele ploegen die disproportioneel vaak voorkomen. Daarnaast werden voor veel teams slechts één of enkele wedstrijden geanalyseerd, wat het bepalen van historische teamkenmerken bemoeilijkt. Toekomstig onderzoek zou er baat bij hebben om volledige seizoenen van specifieke reeksen te verzamelen, zodat de steekproef robuuster en meer representatief wordt.

Verder zijn er belangrijke beperkingen verbonden aan het enquête-onderdeel van dit onderzoek. De online en papieren versies werden verspreid in verschillende periodes: de online enquête aan het einde van het seizoen 2023-2024, de papieren versie bij de start van het seizoen 2024-2025. Dit tijdsverschil kan geleid hebben tot verschillen in contextuele beleving (bijvoorbeeld vermoeidheid vs. hernieuwde motivatie), wat een systematische vertekening van de antwoorden kan veroorzaken. In de toekomst verdient het aanbeveling om beide enquêtevormen gelijktijdig af te nemen.

Daarnaast kon de responsgraad voor beide enquêtevormen niet accuraat worden vastgesteld. Bij de online enquête was het onduidelijk hoeveel wedstrijden de potentiële respondenten effectief nog te spelen hadden. Bij de papieren enquête lag het probleem bij de oncontroleerbare distributie door zaalwachters, waardoor het aantal effectief uitgedeelde formulieren onbekend bleef. Dit bemoeilijkt een betrouwbare inschatting van de responsiviteit en verhindert een volledige vergelijking tussen beide methodes.

Een bijkomende inhoudelijke beperking betreft het prestatieverschil tussen de voorspellende modellen. Wanneer de schepkwaliteit-indicatoren niet beschikbaar waren, daalde de prestatie van de modellen aanzienlijk. De conclusies op basis van deze versies moeten dan ook met de nodige omzichtigheid worden geïnterpreteerd. Toekomstig onderzoek zou kunnen nagaan hoe andere contextuele of gedragsvariabelen de voorspellende waarde van zulke modellen kunnen versterken.

Tenslotte richtte dit onderzoek zich voornamelijk op de statistische analyse van de schepcorner. Er werd onderzocht welke factoren bijdragen aan het succes van een schepcorner en welke impact deze spelhervatting heeft op het wedstrijdverloop en de wedstrijduitslag. De perceptie van de schepcorner door spelers, coaches en supporters werd echter niet meegenomen. Dit vormt een waardevol aanknopingspunt voor vervolgonderzoek dat de belevingsdimensie van de schepcorner binnen het bredere sportieve en sociale kader van het minivoetbal kan verkennen.

Literatuurlijst

- Ahsan, M. M., Mahmud, M. A. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies 2021*, Vol. 9, Page 52, 9(3), 52. <https://doi.org/10.3390/TECHNOLOGIES9030052>
- Armitage, P. (1955). Tests for Linear Trends in Proportions and Frequencies. *Biometrics*, 11(3), 375. <https://doi.org/10.2307/3001775>
- Artino, A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical Teacher*, 36(6), 463–474. <https://doi.org/10.3109/0142159X.2014.889814;WGROU:STRING:PUBLICATION>
- Ball, H. L. (2019). Conducting Online Surveys. *Journal of Human Lactation*, 35(3), 413–417. <https://doi.org/10.1177/0890334419848734>
- Bergstra, J., Ca, J. B., & Ca, Y. B. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13, 281–305. <https://doi.org/10.5555/2188385.2188395>
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- Bransen, L., Robberechts, P., Van, J., & Davis, H. †-J. (2019). Choke or shine? Quantifying soccer players' abilities to perform under mental pressure. *Lirias.Kuleuven.BeL Bransen, P Robberechts, J Van Haaren, J Davis Proceedings of the 13th MIT Sloan Sports Analytics Conference, 2019•lirias.Kuleuven.Be*. <https://lirias.kuleuven.be/retrieve/532444>
- Breiman, L. (2001). Random forests. *Springer*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen, T., He, T., Benesty, M., O.4-2, V. K.-... version, & 2015, undefined. (2025). Xgboost: extreme gradient boosting. *Cran.Ms.Unimelb.Edu.Au*. <https://cran.ms.unimelb.edu.au/web/packages/xgboost/vignettes/xgboost.pdf>
- Clark, T. K., Johnson, A. W., & Stimpson, A. J. (2013). *Going for Three: Predicting the Likelihood of Field Goal Success with Logistic Regression*.
- Couronné, R., Probst, P., & Boulesteix, A. L. (2018). Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics*, 19(1), 1–14. <https://doi.org/10.1186/S12859-018-2264-5/FIGURES/6>
- Crick, J. M., & Crick, D. (2021). Coopetition and sales performance: evidence from non-mainstream sporting clubs. *International Journal of Entrepreneurial Behaviour and Research*, 27(1), 123–147. <https://doi.org/10.1108/IJEBR-05-2020-0273/FULL/PDF>
- de Amorim, L. B. V., Cavalcanti, G. D. C., & Cruz, R. M. O. (2023). The choice of scaling technique matters for classification performance. *Applied Soft Computing*, 133, 109924. <https://doi.org/10.1016/J.ASOC.2022.109924>
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, 44(3), 837. <https://doi.org/10.2307/2531595>

- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Jmlr.OrgJ DemšarJournal of Machine Learning Research, 2006•jmlr.Org, 7*, 1–30. <http://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf>
- Dohmen, T. J. (2008). Do professionals choke under pressure? *Journal of Economic Behavior & Organization, 65*(3–4), 636–653. <https://doi.org/10.1016/J.JEBO.2005.12.004>
- Draugalis, J. L. R., Coons, S. J., & Plaza, C. M. (2008). Best practices for survey research reports: A synopsis for authors and reviewers. *American Journal of Pharmaceutical Education, 72*(1), 11. <https://doi.org/10.5688/aj720111>
- Dunton, G. F. (2017). Ecological Momentary Assessment in Physical Activity Research. *Exercise and Sport Sciences Reviews, 45*(1), 48–54. <https://doi.org/10.1249/JES.0000000000000092>
- Fanning, E. (2005). Formatting a Paper-based Survey Questionnaire: Best Practices. *Practical Assessment, Research, and Evaluation, 10*(1). <https://doi.org/10.7275/S84T-8A63>
- Fawcett, T. (2006). An introduction to ROC analysis. *Elsevier*. <https://www.sciencedirect.com/science/article/pii/S016786550500303X>
- Fricker, R. D., & Schonlau, M. (2002). Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature. *Field Methods, 14*(4), 347–367. <https://doi.org/10.1177/152582202237725>
- Greenhalgh, G. P., Simmons, J. M., Hambrick, M. E., & Greenwell, C. T. (2011). *Spectator Support: Examining the Attributes That Differentiate Niche from Mainstream Sport*. https://www.researchgate.net/profile/Chris-Greenwell-2/publication/288849887_Spectator_support_Examining_the_attributes_that_differentiate_niche_from_mainstream_sport/links/5bb23b54299bf13e6059e30d/Spectator-support-Examining-the-attributes-that-differentiate-niche-from-mainstream-sport.pdf?origin=journalDetail&_tp=eyJwYWdlIjoiam91cm5hbERldGFpbCJ9
- Gregorutti, B., Michel, B., & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *SpringerB Gregorutti, B Michel, P Saint-PierreStatistics and Computing, 2017•Springer, 27*(3), 659–678. <https://doi.org/10.1007/S11222-016-9646-1>
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- Hossin, M., & Sulaiman, M. (2015). A review on evaluation metrics for data classification evaluations. *Academia.Edu*. <https://www.academia.edu/download/37219940/5215ijdkp01.pdf>
- Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering, 17*(3), 299–310. <https://doi.org/10.1109/TKDE.2005.50>
- Hulteen, R. M., Smith, J. J., Morgan, P. J., Barnett, L. M., Hallal, P. C., Colyvas, K., & Lubans, D. R. (2017). Global participation in sport and leisure-time physical activities: A systematic review and meta-analysis. *Preventive Medicine, 95*, 14–25. <https://doi.org/10.1016/J.YPMED.2016.11.027>
- Jordet, G., Hartman, E., Visscher, C., & Lemmink, K. A. P. M. (2007). Kicks from the penalty mark in soccer: The roles of stress, skill, and fatigue for kick outcomes. *Journal of Sports Sciences, 25*(2), 121–129. <https://doi.org/10.1080/02640410600624020>

- Koseler, K., & Stephan, M. (2017). Machine Learning Applications in Baseball: A Systematic Literature Review. *Applied Artificial Intelligence*, 31(9–10), 745–763.
<https://doi.org/10.1080/08839514.2018.1442991>;JOURNAL:JOURNAL:UAAI20;REQUESTEDJOURNAL:JOURNAL:UAAI20;WGROU:STRING:PUBLICATION
- Krosnick, J. A. (2017). Questionnaire design. In *The Palgrave handbook of survey research* (pp. 439–455).
- Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47(260), 583–621.
<https://doi.org/10.1080/01621459.1952.10483441>;WGROU:STRING:PUBLICATION
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(11), 1–13. <https://doi.org/10.18637/JSS.V036.I11>
- Landis, J. R., & Koch, G. G. (1977). An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics*, 33(2), 363.
<https://doi.org/10.2307/2529786>
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, 96(1), 3–14.
<https://doi.org/10.1080/00220670209598786>
- Seger, C. (2018). An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. *DEGREE PROJECT TECHNOLOGY*.
<https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-237426>
- Senaviratna, N., and, T. C.-A. J. of P., & 2019, undefined. (2019). Diagnosing multicollinearity of logistic regression model. *Researchgate.NetN Senaviratna, T CoorayAsian Journal of Probability and Statistics*, 2019•researchgate.Net. https://www.researchgate.net/profile/Namr-Senaviratna/publication/336213669_Diagnosing_Multicollinearity_of_Logistic_Regression_Model/links/5d970669a6fdccfd0e74882e/Diagnosing-Multicollinearity-of-Logistic-Regression-Model.pdf
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annualreviews.Org*, 4, 1–32. <https://doi.org/10.1146/ANNUREV.CLINPSY.3.022806.091415>
- Siva Durga Prasad Nayak, M., & Narayan, K. A. (2019, May). *Strengths and weaknesses of online surveys*. https://www.researchgate.net/profile/Mudavath-Nayak/publication/333207786_Strengths_and_Weakness_of_Online_Surveys/links/61176e5a0c2bfa282a42253b/Strengths-and-Weakness-of-Online-Surveys.pdf
- Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2), 130. <https://doi.org/10.11919/J.ISSN.1002-0829.215044>
- Toloşi, L., Bioinformatics, T. L.-, & 2011, undefined. (2011). Classification with correlated features: unreliability of feature ranking and solutions. *Academic.Oup.ComL Toloşi, T LengauerBioinformatics*, 2011•academic.Oup.Com, 27(14), 1986–1994.
<https://doi.org/10.1093/bioinformatics/btr300>

- Turocy, P. S. (2002). Survey Research in Athletic Training: The Scientific Method of Development and Implementation. *Journal of Athletic Training*, 37(4 suppl), S-174.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC164421/>
- Urbanski, M. T. (2016). *Tree boosting with xgboost-why does xgboost win" every" machine learning competition?* https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2433761/16128_FULLTEXT.pdf
- Wright, K. B. (2005). Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, 10(3). <https://doi.org/10.1111/J.1083-6101.2005.TB00259.X/4614509>
- Wu, M. J., Zhao, K., & Fils-Aime, F. (2022). Response rates of online surveys in published research: A meta-analysis. *Computers in Human Behavior Reports*, 7, 100206.
<https://doi.org/10.1016/J.CHBR.2022.100206>
- Yang, L., & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295–316.
<https://doi.org/10.1016/J.NEUCOM.2020.07.061>
- Zec, S., Soriani, N., Comoretto, R., & Baldi, I. (2017). High Agreement and High Prevalence: The Paradox of Cohen's Kappa. *The Open Nursing Journal*, 11(1), 211.
<https://doi.org/10.2174/1874434601711010211>

Appendix

A.1 Papieren enquête versie 1

Enquête Cornerbal

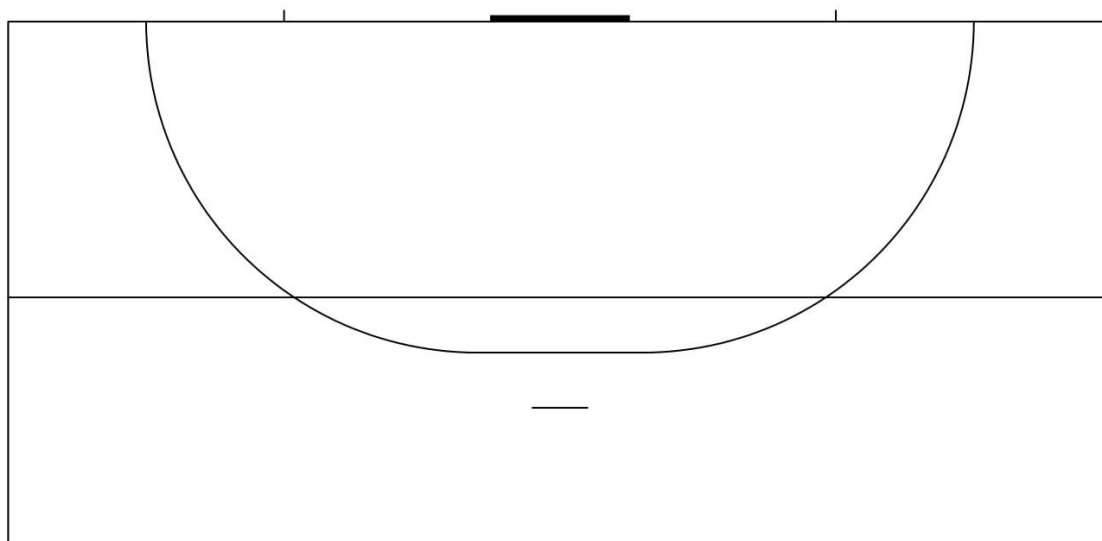
Datum:

Reeks:

	Thuisploeg	Uitploeg
Clubnaam		
Rugnummer Schepper		
Rugnummer Kopper		
Positie in klassement		
Ervaring van koper/schepper/duo geef een score van 1-5		
Schepper		
Kopper		
Duo		

OMCIRKEL HET RELEVANTE ANTWOORD

Schepcorner voor: **THUISPLOEG** **UITPLOEG**
Kwart: **1** **2** **3** **4**
Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**
Score voor schepcorner:-.....
Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan
met een **cirkel** de positie van de koper voor de schepper schept
met een **driehoek** de positie van de koper wanneer hij de bal kopt.



schepcorner gescoord: **JA** **NEE**

Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

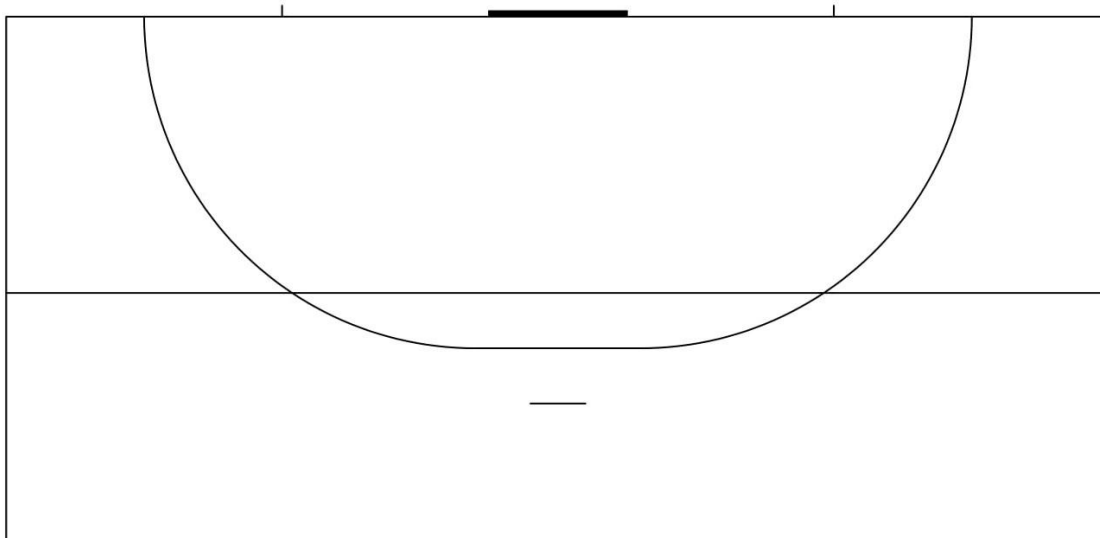
Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan

met een **cirkel** de positie van de koper voor de schepper schept

met een **driehoek** de positie van de koper wanneer hij de bal kopt.



Schepcorner gescoord: **JA** **NEE**

Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

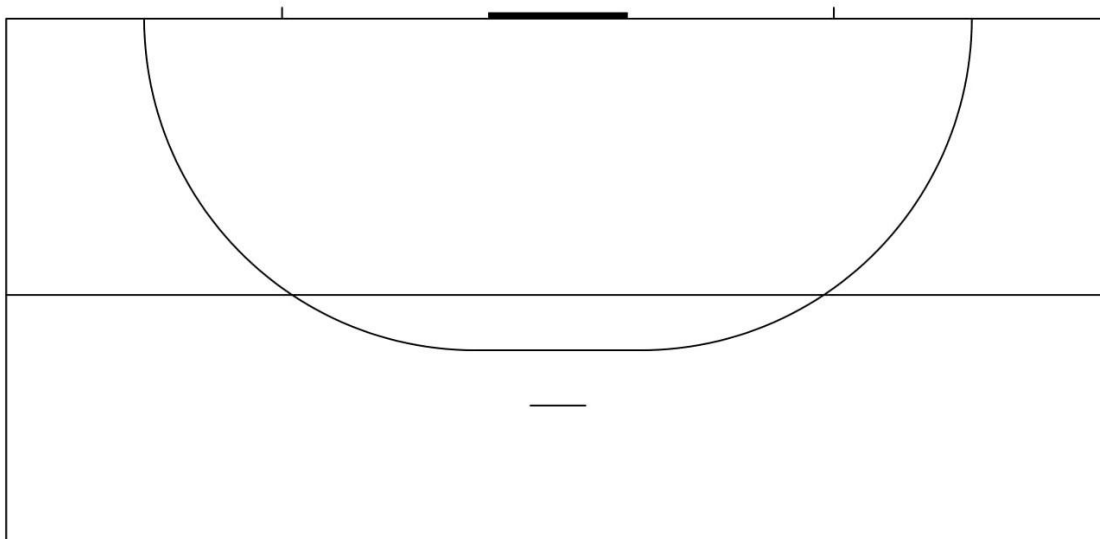
Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan

met een **cirkel** de positie van de koper voor de schepper schept

met een **driehoek** de positie van de koper wanneer hij de bal kopt.



Schepcorner gescoord: **JA** **NEE**

Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

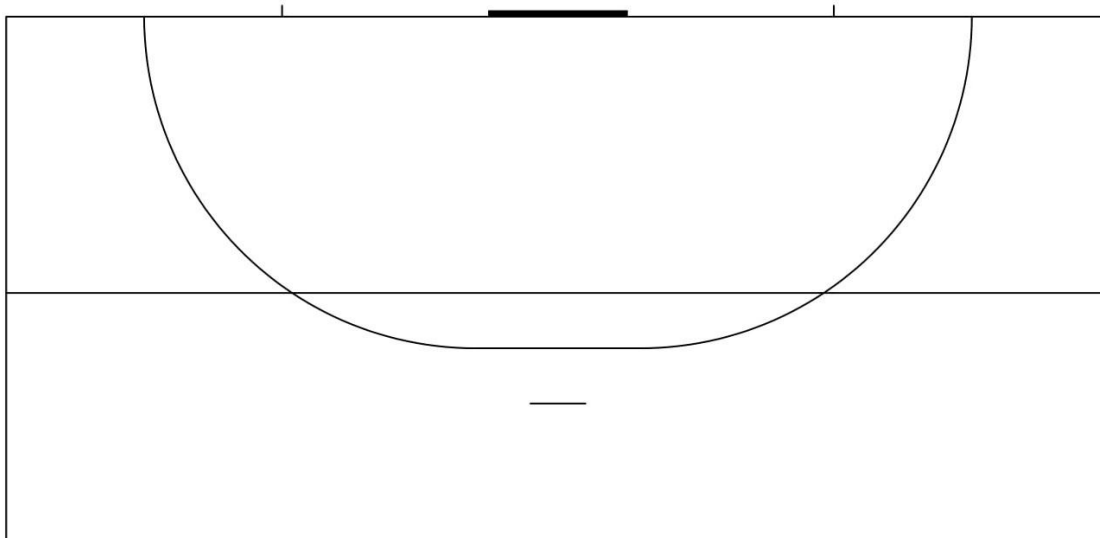
Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan

met een **cirkel** de positie van de koper voor de schepper schept

met een **driehoek** de positie van de koper wanneer hij de bal kopt.



Schepcorner gescoord: **JA** **NEE**

Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

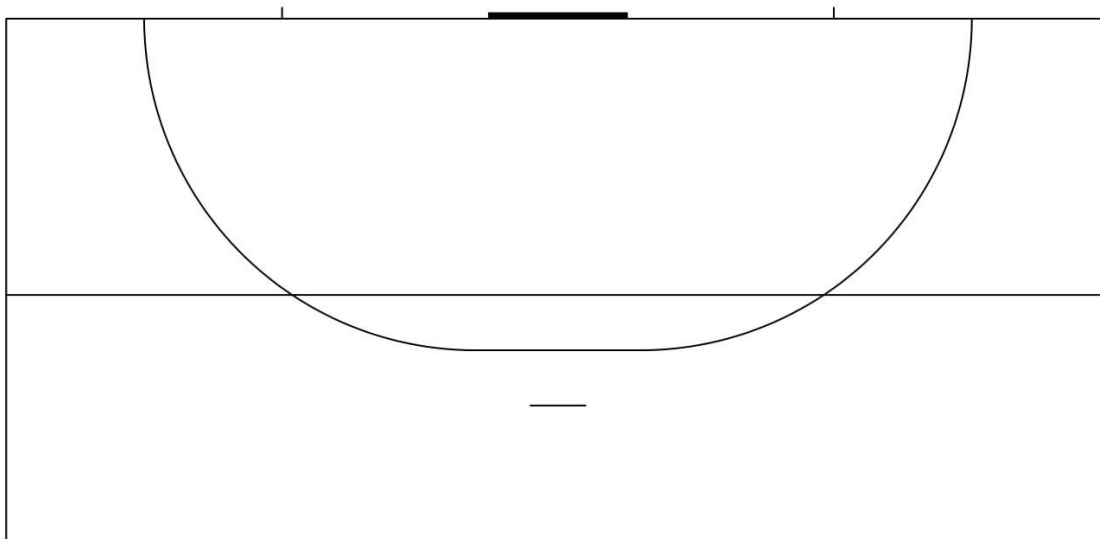
Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan

met een **cirkel** de positie van de koper voor de schepper schept

met een **driehoek** de positie van de koper wanneer hij de bal kopt.



Schepcorner gescoord: **JA** **NEE**

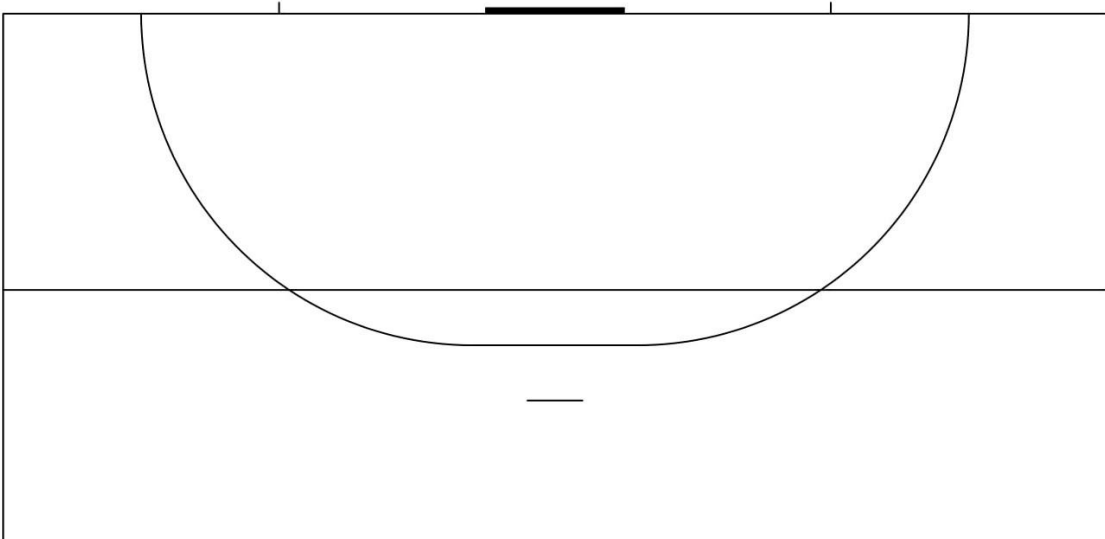
Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan
met een **cirkel** de positie van de koper voor de schepper schept
met een **driehoek** de positie van de koper wanneer hij de bal kopt.



Schepcorner gescoord: **JA** **NEE**

Uiteindelijke uitslag									
	Thuis				Uit				
Goals									
Corners (op scorebord, niet totaal aantal schepcorners)	0	1	2	3	0	1	2	3	

A.2 Papieren enquête versie 2

Enquête Cornerbal

Datum:

Reeks:

	Thuisploeg	Uitploeg
Clubnaam		
Rugnummer Schepper		
Rugnummer Kopper		
Positie in klassement		
Ervaring van koper/schepper/duo geef een score van 1-5		
Schepper		
Kopper		
Duo		
Uiteindelijke uitslag (invullen op het einde van de match)		
	Thuis	Uit
Goals		
Corners (op scorebord, niet totaal aantal schepcorners)	0 1 2 3	0 1 2 3

Vergeet op de volgende pagina's zeker niet de info per schepcorner op te schrijven!

OMCIRKEL HET RELEVANTE ANTWOORD

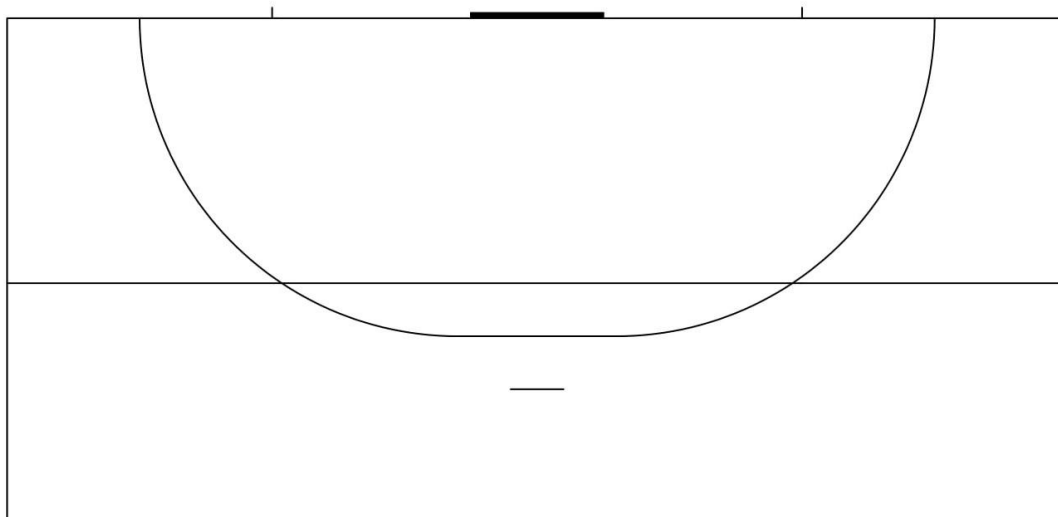
Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan
met een **cirkel** de positie van de koper voor de schepper schept
met een **driehoek** de positie van de koper wanneer hij de bal kopt.



schepcorner gescoord: **JA** **NEE**

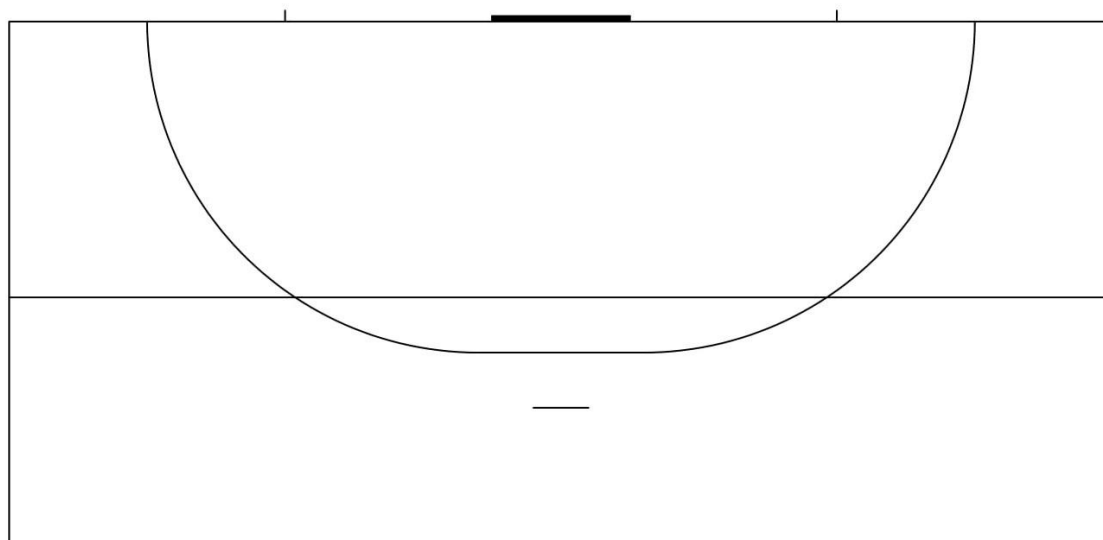
Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan
met een **cirkel** de positie van de koper voor de schepper schept
met een **driehoek** de positie van de koper wanneer hij de bal kopt.



Schepcorner gescoord: **JA** **NEE**

Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

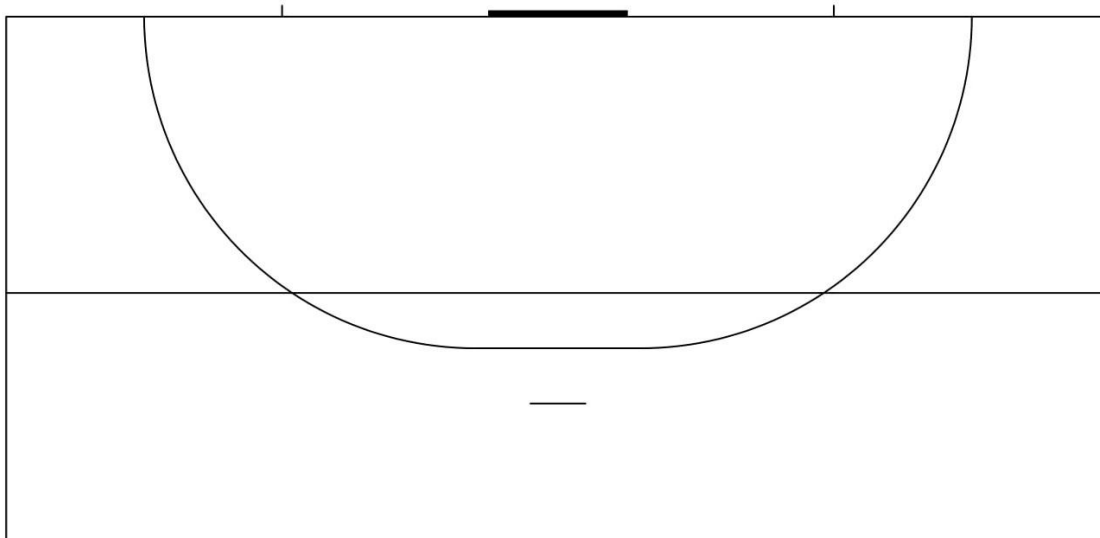
Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan

met een **cirkel** de positie van de koper voor de schepper schept

met een **driehoek** de positie van de koper wanneer hij de bal kopt.



Schepcorner gescoord: **JA** **NEE**

Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

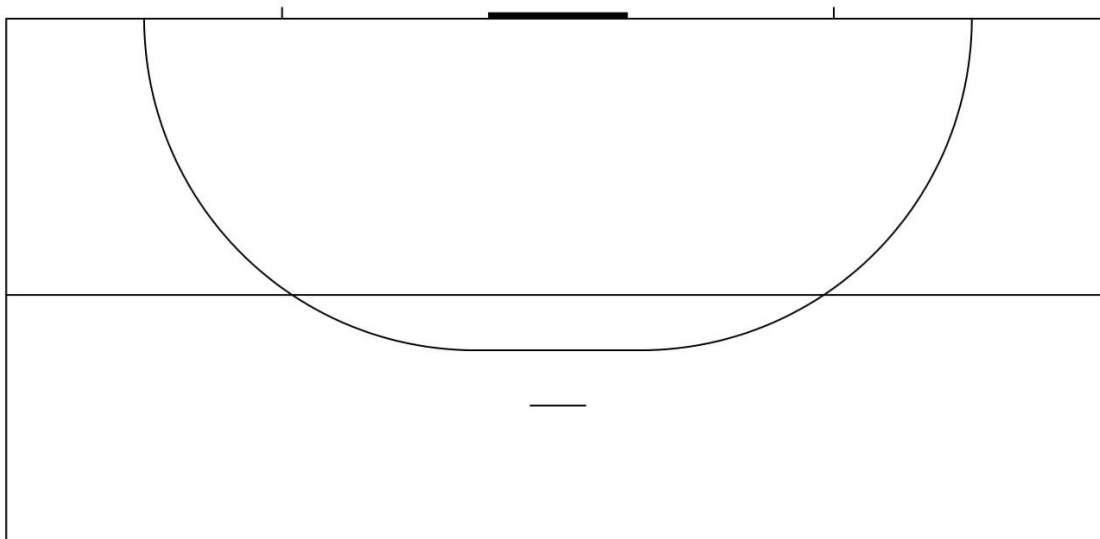
Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan

met een **cirkel** de positie van de koper voor de schepper schept

met een **driehoek** de positie van de koper wanneer hij de bal kopt.



Schepcorner gescoord: **JA** **NEE**

Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

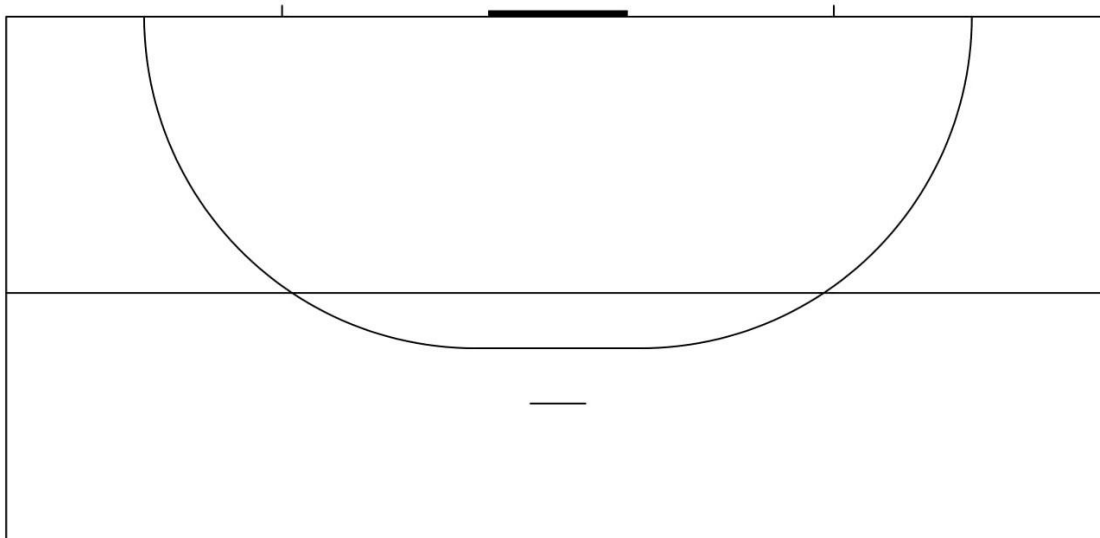
Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan

met een **cirkel** de positie van de koper voor de schepper schept

met een **driehoek** de positie van de koper wanneer hij de bal kopt.



Schepcorner gescoord: **JA** **NEE**

Schepcorner voor: **THUISPLOEG** **UITPLOEG**

Kwart: **1** **2** **3** **4**

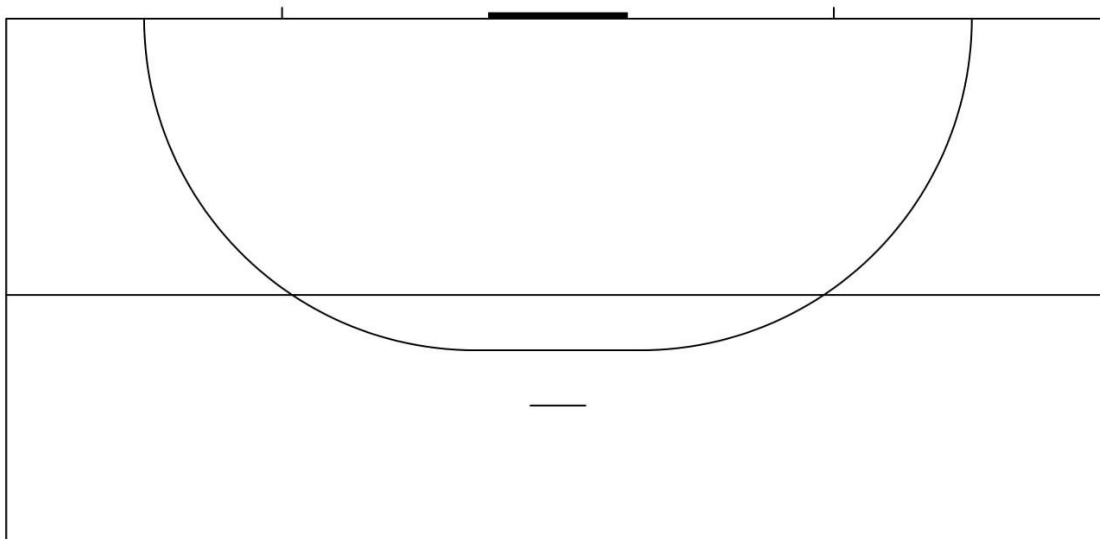
Kwam iemand van de bank: **SCHEPPER** **KOPPER** **BEIDEN SPEELDEN**

Score voor schepcorner:-.....

Duid op onderstaande afbeelding met een **kruis** de schepper zijn positie aan

met een **cirkel** de positie van de koper voor de schepper schept

met een **driehoek** de positie van de koper wanneer hij de bal kopt.



Schepcorner gescoord: **JA** **NEE**

A.3 Hyperparameters Random Forest

Hyperparameter	Mogelijke waarden
n_estimators	100, 200, 300
max_depth	None, 10, 20, 30
min_samples_split	2, 5, 10
min_samples_leaf	1, 2, 4
max_features	sqrt, log2
class_weight	None, balanced

Tabel 20: Hyperparameters Random Forest

Model	Geselecteerde hyperparameters
Random Forest	n_estimators=100, max_depth=None, min_samples_split=10, min_samples_leaf=2, max_features='sqrt', class_weight='balanced'
RF permutation importance	n_estimators=100, max_depth=10, min_samples_split=10, min_samples_leaf=2, max_features='sqrt', class_weight=None
RF boruta	n_estimators=200, max_depth=None, min_samples_split=2, min_samples_leaf=4, max_features='sqrt', class_weight='balanced'
RF zonder schepinfo	n_estimators=100, max_depth=None, min_samples_split=10, min_samples_leaf=4, max_features='sqrt', class_weight=None
RF permutation importance zonder schepinfo	n_estimators=100, max_depth=None, min_samples_split=2, min_samples_leaf=4, max_features='sqrt', class_weight=None
RF boruta zonder schepinfo	n_estimators=200, max_depth=None, min_samples_split=10, min_samples_leaf=4, max_features='sqrt', class_weight=None

Tabel 21: Geselecteerde hyperparameters voor Random Forest modellen

A.4 Hyperparameters XGBoost

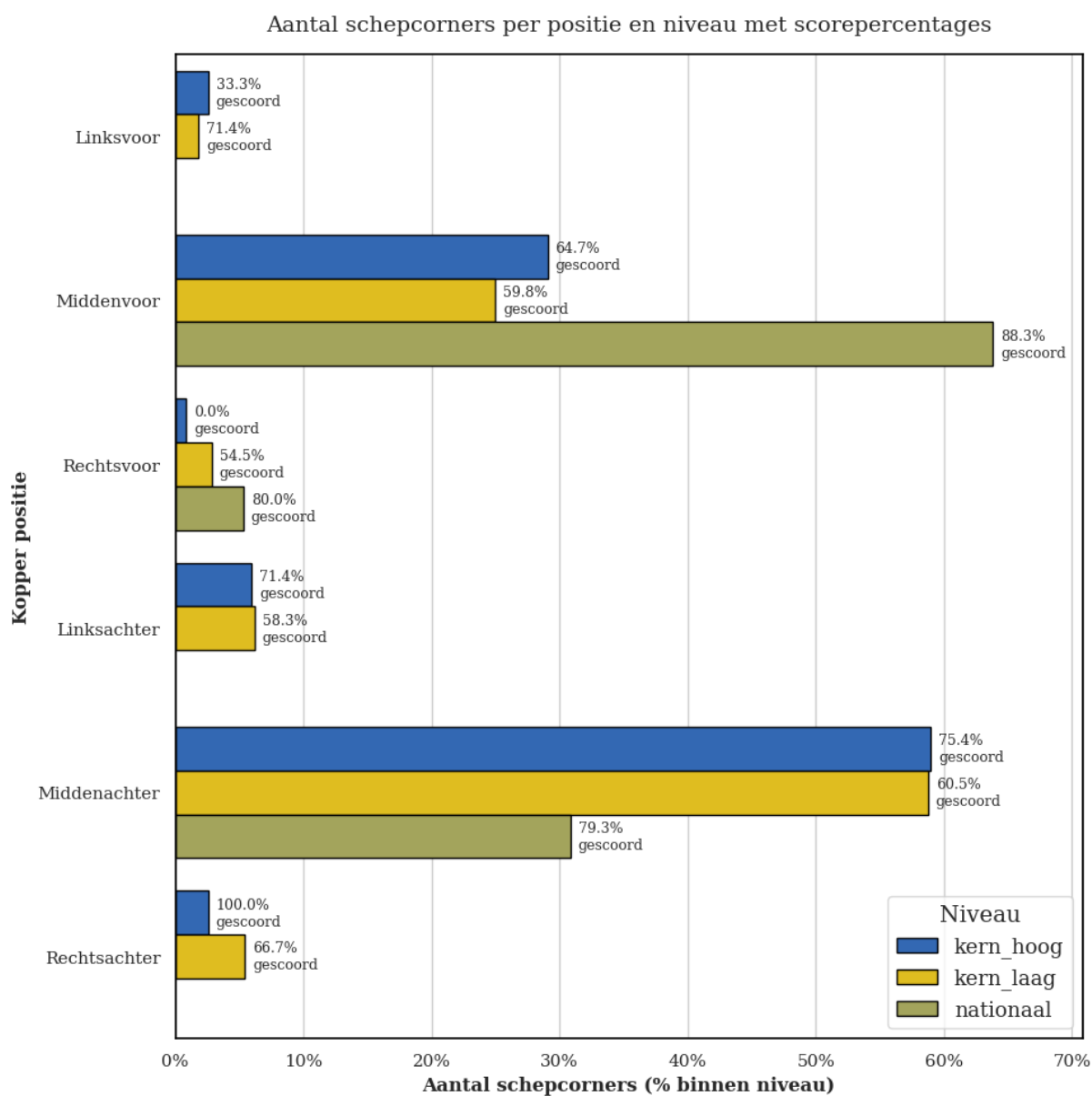
Hyperparameter	Mogelijke waarden
n_estimators	100, 200
max_depth	3, 6, 10
learning_rate	0.01, 0.1, 0.2
subsample	0.8, 1
colsample_bytree	0.8, 1
scale_pos_weight	0.5, 1

Tabel 22: Hyperparameters XGBoost

Model	Geselecteerde hyperparameters
XGBoost	n_estimators=100, max_depth=3, learning_rate=0.2, subsample=0.8, colsample_bytree=0.8, scale_pos_weight=1
XGBoost permutation importance	n_estimators=100, max_depth=3, learning_rate=0.1, subsample=1, colsample_bytree=0.8, scale_pos_weight=1
XGBoost boruta	n_estimators=200, max_depth=3, learning_rate=0.1, subsample=1, colsample_bytree=1, scale_pos_weight=1
XGBoost zonder schepinfo	n_estimators=100, max_depth=3, learning_rate=0.01, subsample=0.8, colsample_bytree=0.8, scale_pos_weight=1
XGBoost permutation importance zonder schepinfo	n_estimators=200, max_depth=3, learning_rate=0.01, subsample=0.8, colsample_bytree=0.8, scale_pos_weight=1
XGBoost boruta zonder schepinfo	n_estimators=200, max_depth=10, learning_rate=0.01, subsample=0.8, colsample_bytree=0.8, scale_pos_weight=0.5

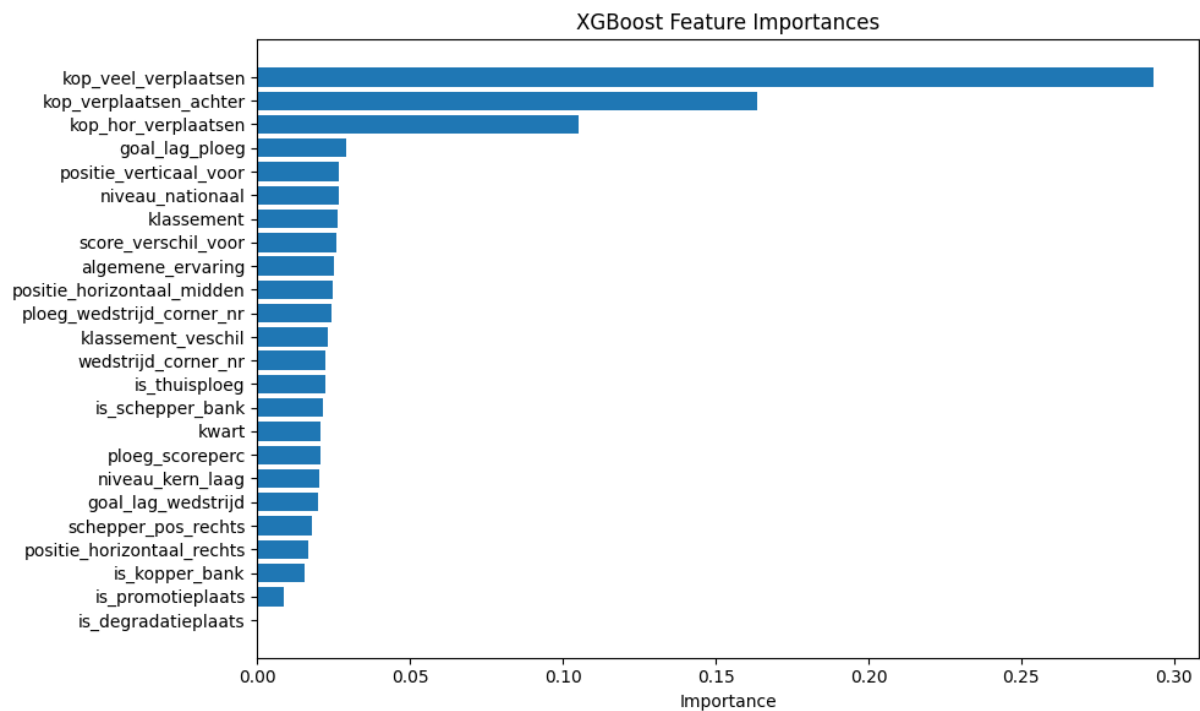
Tabel 23: Geselecteerde hyperparameters voor XGBoost modellen

A.5 Posities per niveau

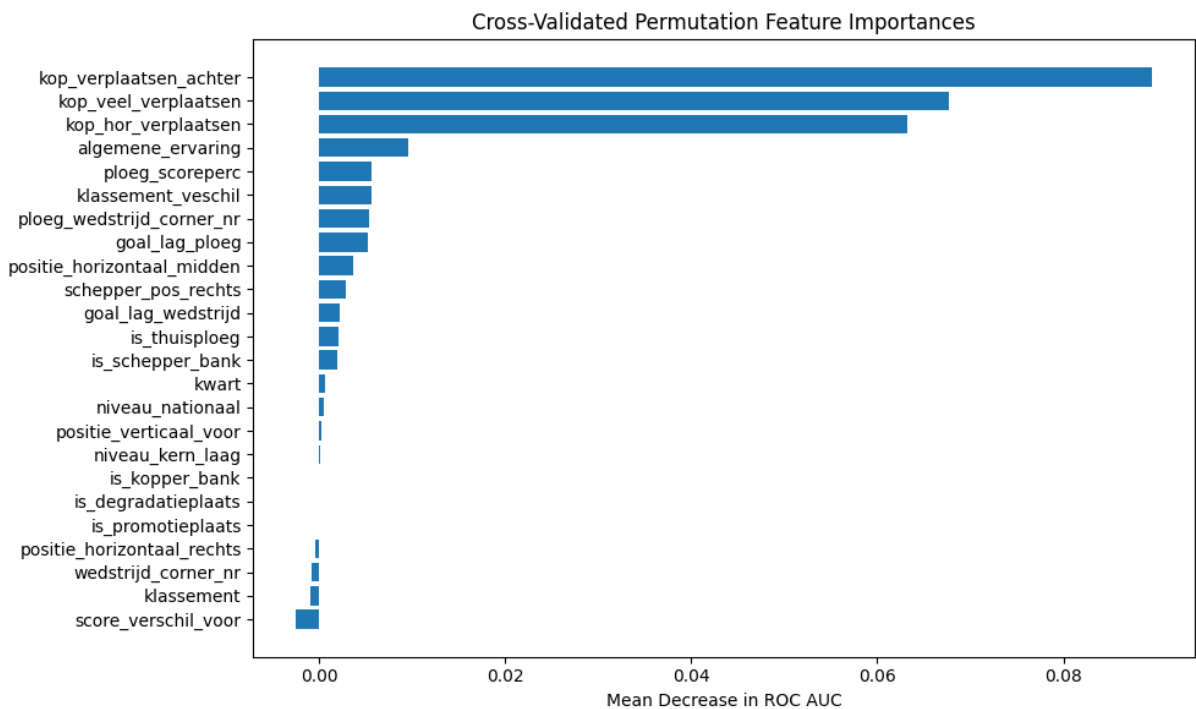


Figuur 13: Posities per niveau

A.6 Visualisaties importance XGBoost



Figuur 14: XGBoost feature importance



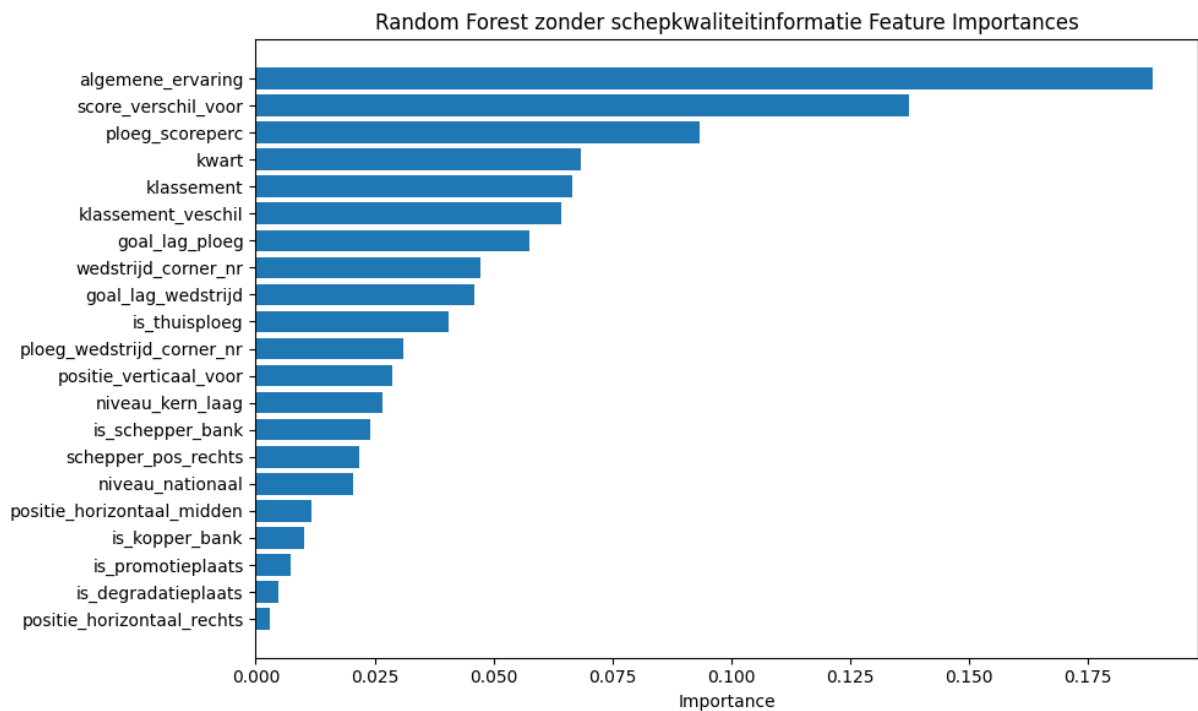
Figuur 15: XGBoost permutation importance

A.7 Confusion Matrix

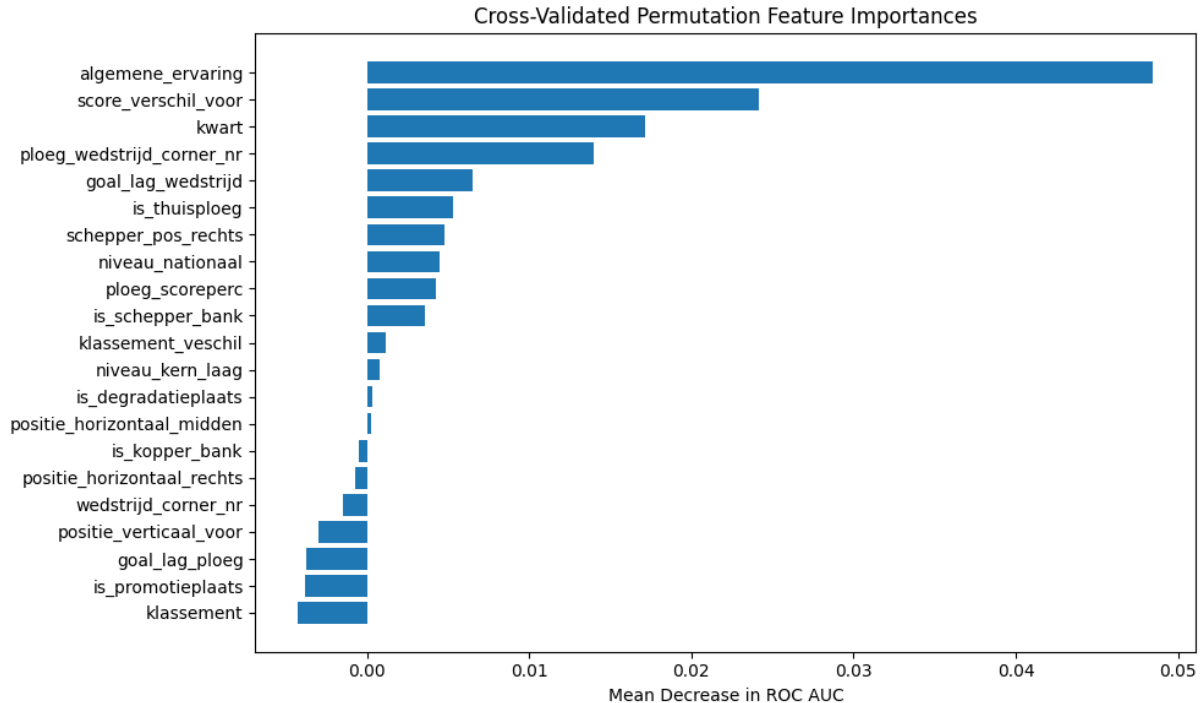
Model	True Positive	False Positive	True Negative	False Negative
Baseline LR	79	21	21	0
Baseline LR zonder schepinfo	65	33	9	14
Random Forest	76	17	25	3
RF permutation importance	77	18	24	2
RF boruta	76	15	27	3
RF zonder schepinfo	75	38	4	4
RF permutation importance zonder schepinfo	72	34	8	7
RF boruta zonder schepinfo	63	37	5	16
XGBoost	77	20	22	2
XGBoost permutation Importance	77	17	25	2
XGBoost boruta	77	22	20	2
XGBoost zonder schepinfo	78	42	0	1
XGBoost permutation importance zonder schepinfo	72	35	7	7
XGBoost boruta zonder schepinfo	28	8	34	51

Tabel 24: Voorspellingsresultaten per model

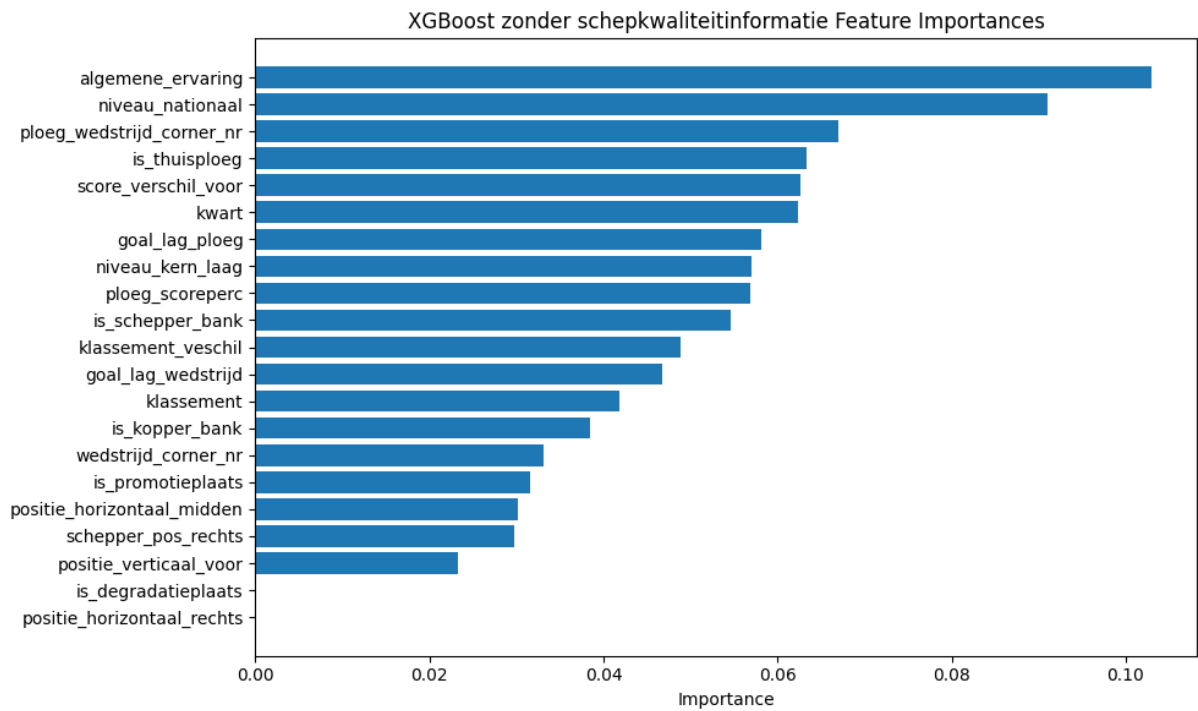
A.8 Visualisaties importance zonder schepkwaliteit



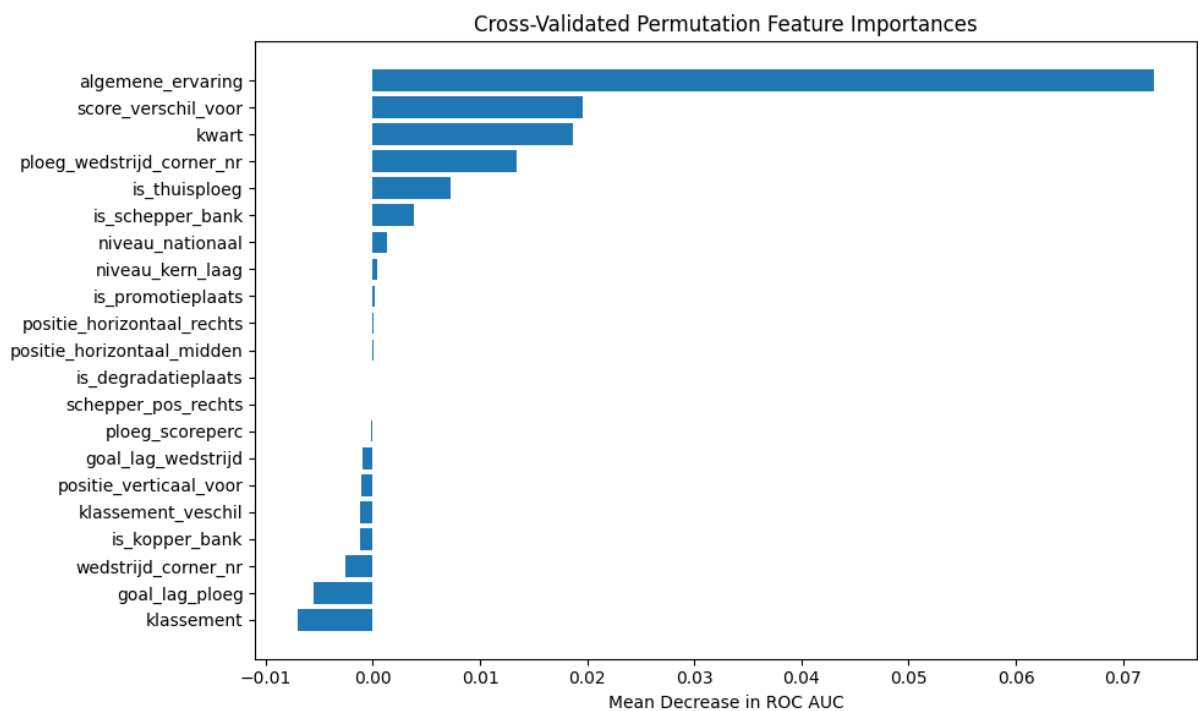
Figuur 16: RF zonder schepkwaliteit feature importance



Figuur 17: RF zonder schepkwaliteit permutation importance



Figuur 18: XGB zonder schepkwaliteit feature importance



Figuur 19: XGB zonder schepkwaliteit permutation importance