

Leseprobe

Digitale Kaderplanung und Datenscouting (IST)

Studienheft

Datenscouting

Autor

Boris Notzon (Diplom-Sportmanager)

1. Evolution des Datenscouting und Abgrenzung der wesentlichen Kennzeichnungen

Positionsdatenerfassung

Die Positionsdaten beschreiben den Aufenthaltsort eines Spielers während des Spiels. Dabei wird das Spielfeld in ein Koordinatensystem mit X- und Y-Achse verwandelt. Den Spielern kann dann zu jedem Aufenthaltsort ein X- und ein Y-Wert zugeordnet werden. Die Werte können mithilfe verschiedener Technologien, wie z. B. spezielle Kamerasysteme oder Sensoren unter oder am Trikot, erfasst werden. Diese Werte werden in Echtzeit generiert und können so auch schnellstmöglich genutzt werden. Beispielsweise kann so ein Spielzug, der zu einem Tor oder Gegentor führt, auf Basis der Positionsdaten innerhalb weniger Sekunden ausgewertet werden, sowohl für die eigene Mannschaft als auch für den Gegner. Die Positionsdaten helfen außerdem dabei, die Leistungen eines Spielers und der Mannschaft zu bewerten und das auf taktischer, technischer und physischer Ebene. Ziel ist es, dadurch Erkenntnisse für Spiel und Training zu erhalten und sich dadurch einen Vorteil zu erarbeiten. Ein weiteres Ziel ist die Erkennung von Spielmustern im eigenen Spiel und dem des Gegners (Memmert et al., 2014).

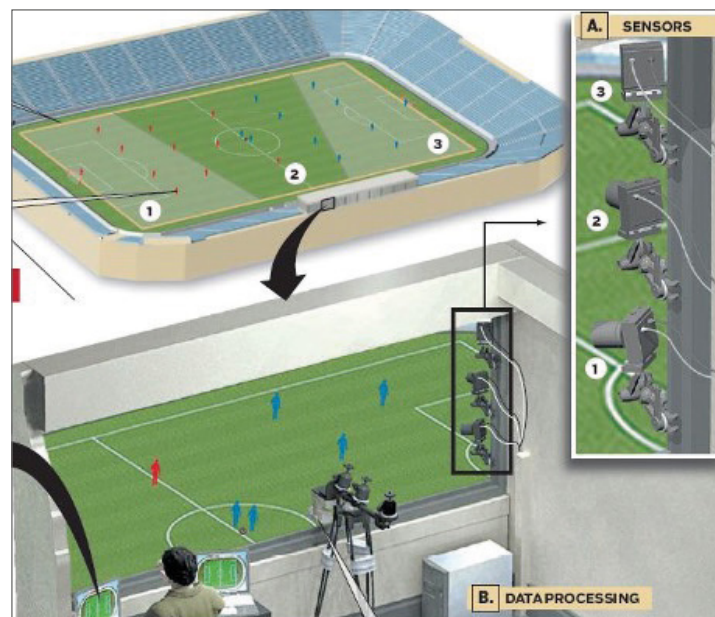


Abb. 6 Tracking-Systematik im Profifußball
(Matzat, 2010)

1. Evolution des Datenscouting und Abgrenzung der wesentlichen Kennzeichnungen

Das französische Unternehmen Amisco entwickelte 1996 das erste Tracking-System, welches mit Wärmebildkameras ausgestattet war. Das Ziel war die Erfassung von Positionsdaten aller 22 Spieler, anhand der Lokalisation der Spieler auf dem Feld. Die physischen Datenauswertungen dienten zur Analyse des Bewegungsverhalten der Spieler und sollte erstmalig datengestützte Rückschlüsse zum präzisen mannschaftstaktischen Verhalten geben. Im Laufe der Entwicklung des Marktes der Anbieter von Tracking-Systemen kam das englische Unternehmen Prozone mit einem ähnlichen System auf den Markt und 2011 fusionierten die beiden Unternehmen Amisco und Prozone. Sie bedienten lange als Marktführer viele internationale Ligen und Verbände mit ihrem System zur physischen Datenerfassung. Heute sind mehrere Anbieter solcher IT-Lösungen etabliert und es kommen verschiedene Technologien zum Einsatz. Beschleuniger der Entwicklung moderner Trackingmethoden waren die Fortschritte, welche in den vergangenen Jahren im Bereich der Softwareentwicklung gemacht wurden. Elementar wichtig waren die Durchbrüche im Bereich der Bildverarbeitung. Erst diese ermöglichten es, aus mehreren Kameraperspektiven zuverlässig einzelne Spieler und ihre Bewegungen zu extrahieren. Heute dienen Algorithmen zur Erkennung von Personen oder Gesichtern in bewegten Bildern. Darüber hinaus bilden komplexe Algorithmen die Basis zur Analyse der gesammelten Informationen. Ohne eine geschickte Verarbeitung und Auswertung der Datenmengen ist ihr sportlicher Mehrwert sehr gering. Mit den richtigen Modellen hingegen ist es möglich, die computergestützte Spielanalyse weitestgehend zu automatisieren, sodass komplexe Berechnungen bereits live während eines Spiels zur Verfügung stehen (Memmert et al., 2014). Im Jahr 2018 beschloss die FIFA zudem die Erlaubnis der Nutzung von sogenannten ‚wearables‘, also das Tragen von technischen Hilfsmitteln, in Form von Sensoren am Körper, im Wettkampf. Die Erhebung von Positionsdaten erfolgt über ein videobasiertes Trackingsystem, welches Informationen über die auf dem Spielfeld befindlichen Objekte zur Verfügung stellt. Jeder Frame einer Aufzeichnung besteht in seine Grundstruktur aus Metadaten, Personendaten und Balldaten (siehe Abbildung).

Amisco



© IST-Studieninstitut

1. Evolution des Datenscouting und Abgrenzung der wesentlichen Kennzeichnungen

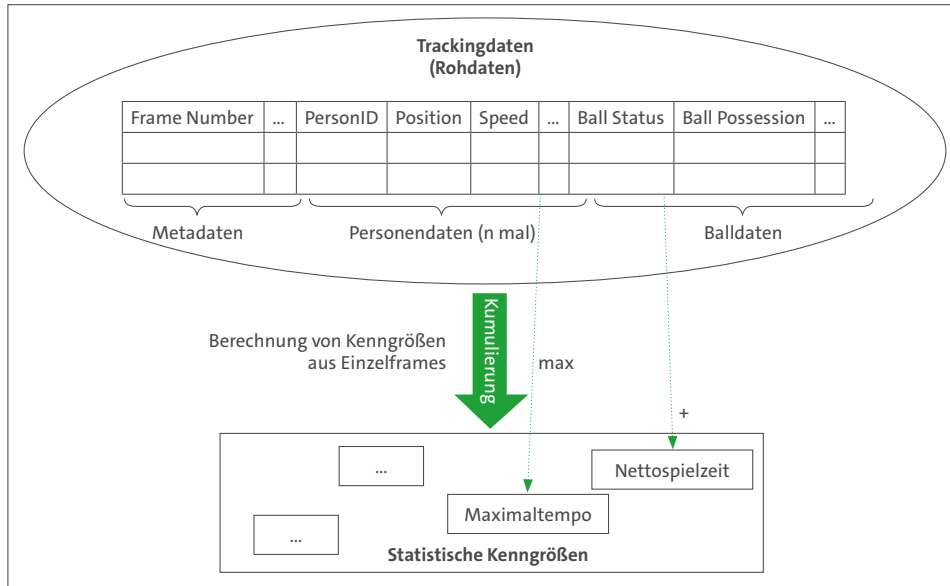


Abb. 7 Erfassung von Trackingdaten nach DFL-Definitionskatalog (Bundesliga, 2013)

Personendaten Die Personendaten beinhalten Personen-ID, Position und Tempo für jede am Spiel beteiligte Person (inklusive Schiedsrichter). Die Balldaten enthalten Angaben über den Spielzustand (Ball im Spiel, Ball nicht im Spiel) und den Ballbesitz (Heimmannschaft oder Auswärtsmannschaft). Statistische Kenngrößen werden auf Basis aller zu einem Spiel bereitgestellten Frames berechnet, beispielsweise ergibt sich das Maximaltempo eines Spielers als Maximum aller Geschwindigkeitsangaben zu diesem Spieler. Unterhalb aufgeführte Übersicht beschreibt im Detail, wie auf Basis dieser Rohdaten Distanzen, Gesamtlafstrecken, Sprints oder Laufrichtungen berechnet werden. Bei der Verwendung solcher Kenngrößen sollte dem Datenscout ein exaktes Verständnis geläufig sein, um unter anderem Parameter wie Maximaltempo, Sprint und Durchschnittstempo einordnen zu können.

1. Evolution des Datenscouting und Abgrenzung der wesentlichen Kennzeichnungen

Offizielle Bezeichnung	Beschreibung
Laufdistanz	Die Laufdistanz eines Spielers zwischen zwei Frames ist gegeben durch den euklidischen Abstand der Positionen in xy-Ebene.
Gesamtlaufdistanz	Die Gesamtlaufdistanz eines Spielers oder einer Mannschaft ist gegeben durch die Summe aller Laufdistanzen in allen bereitgestellten Frames.
Durchschnittstempo	Das Durchschnittstempo ist gegeben durch den Durchschnittswert aller Tempoangaben für einen Spieler.
Maximaltempo	Das Maximaltempo ist gegeben über den Maximalwert aller Tempoangaben für einen Spieler.
Tempoläufe	Ein Tempolauf liegt vor, wenn ein Spieler mehr als zwei Sekunden mehr als 4,0 m/s läuft und während dieser Zeit zwischen 5,0 m/s und 6,3 m/s läuft, wobei zwischen dem ersten Erreichen und dem letzten Unterschreiten der 5,0 m/s Schwelle mindestens 1 Sekunde liegt. Der Tempolauf ist zeitlich auf die Zeitspanne begrenzt, in der die 5,0 m/s-Schwelle das erste Mal überschritten und das letzte Mal unterschritten wird. Ein neuer Tempolauf kann erst dann vorliegen, wenn die Schwelle von 4 m/s unterschritten wird.
Sprints	Ein Sprint liegt vor, wenn ein Spieler mehr als zwei Sekunden mindestens 4,0 m/s läuft und während dieser Zeit mindestens 6,3 m/s läuft, wobei zwischen dem ersten Erreichen und dem letzten Unterschreiten der 6,3 m/s-Schwelle mindestens 1 Sekunde liegt. Der Sprint ist zeitlich auf die Zeitspanne begrenzt, in der die 6,3 m/s-Schwelle das erste Mal erreicht wird und das letzte Mal unterschritten wird. Ein neuer Sprint kann erst dann vorliegen, wenn die Schwelle von 4,0 m/s unterschritten wird.
Intensive Läufe	Die Anzahl der intensiven Läufe ist gegeben über die Anzahl aller Tempoläufe und Sprints.
Offensive Läufe	Ein offensiver Lauf liegt vor, wenn ein Spieler in den Spielfeldbereich läuft, der weniger als 25 Meter von der gegnerischen Torauslinie entfernt ist. Ein Lauf kann immer nur als ein Lauf gewertet werden.
Läufe in den gegnerischen Strafraum	Ein Lauf in den gegnerischen Strafraum liegt vor, wenn ein Spieler in den gegnerischen Strafraum läuft. Ein Lauf kann immer nur als ein Lauf gewertet werden.
Distanz Tempoläufe	Ist die Distanz, die ein Spieler im Rahmen von Tempoläufen absolviert hat.
Distanz Sprints	Ist die Distanz, die der Spieler im Rahmen von Sprints absolviert hat.
Laufrichtung	Die Laufrichtung eines Spielers bei Sprints und Tempoläufen ergibt sich aus seiner Ortskoordination zum Beginn und Ende eines solchen Laufes. vorwärts: die x-Koordinate liegt zum Ende des Laufes näher am gegnerischen Tor als Beginn des Laufes rückwärts: in allen anderen Fällen
Tempobereiche	1: <11 km/h 2: 11–14 km/h 3: 14–17 km/h 4: 17–21 km/h 5: 21–24 km/h 6: >24 km/h
Zeit im Tempobereich n	Anzahl der Frames eines Spielers, in denen das Tempo im Tempobereich n liegt multipliziert mit dem Kehrwert der Messfrequenz (zurzeit 40 ms).
Laufdistanz im Tempobereich n	Summe aller Laufdistanzen eines Spielers zwischen zwei aufeinanderfolgenden Frames, bei denen das Tempo des zweiten Frames im Tempobereich n liegen.
Nettospielzeit	Anzahl der Frames, in denen der Ball im Spiel ist, multipliziert mit dem Kehrwert der Messfrequenz (zurzeit 40 ms).
Ballbesitz (Verhältnis)	Verhältnis der prozentualen Anteile am zeitlichen Ballbesitz von Heim- und Auswärtsmannschaft (z. B. 55 % : 45 %).

Tab. 3 Beschreibung und Definition der einzelnen Tracking-Werte nach DFL-Definitionskatalog (Bundesliga, 2013)



1.6 Big Data und Key Performance Index

Big Data Welche Möglichkeiten ergeben sich aus Datenanalysen, die nach Festlegung aller Definitionskriterien und Definitionskataloge erfasst werden? Hier fällt seit einigen Jahren immer wieder der Ausdruck **Big Data** im Zusammenhang mit Datenanalysen im Profifußball. Der Begriff ‚Big Data‘ ist recht vage definiert und wandelt sich dabei auch kontinuierlich. Als medial sehr häufig verwendetes Synonym umfasst Big Data verschiedene Technologien wie Cloud Computing, Parallel Computing, Data Mining oder Business Intelligence. Grundsätzlich werden Datensätze, die so gigantisch groß sind, dass sie mit herkömmlichen Datenverarbeitungssystemen nicht bewältigt werden können, als Big Data bezeichnet. Das Speichern großer Datenmengen oder der Zugriff darauf zu Analysezwecken ist nicht neu. Aber erst in den frühen 2000er-Jahren gewann der Begriff an Bedeutung, als der Branchenanalytiker Doug Laney die heute weitestgehend anerkannte Definition von Big Data in seinem 3-V-Modell formulierte (Laney, 2001).

Das **3-V-Modell** bezieht sich auf folgende Attribute:

Volume: Der riesige Umfang der generierten und gespeicherten Daten im Petabereich, Exabereich und Zetabytebereich.

Velocity: Die enorme Datenrate und ausreichende Geschwindigkeit bei der Datenverarbeitung und -analyse.

Variety: Die Vielzahl unterschiedlicher Datenformate, -typen und -quellen.

Um aus Big-Data-Werten neue werthaltige Erkenntnisse zu schöpfen, werden typischerweise folgende Phasen durchlaufen.

Datengenerierung → Datenerfassung → Datenspeicherung → Datenanalyse
(Gruber, 2017)

1. Evolution des Datenscouting und Abgrenzung der wesentlichen Kennzeichnungen

Der folgerichtig nächste Schritt nach erfolgreicher Erfassung von Daten ist die Datenanalyse im Bereich der dynamisch-taktischen Auswertung von Positionsdaten und der Etablierung neuer qualitativer Key-Performance-Indikatoren (KPI) in diesem Bereich. In Bezug auf die taktische Komponente des Spiels zeigen beispielsweise die Datensätze der Firma Impect spieltaktische Analysen, welche unter dem Oberbegriff ‚Packing‘ bekannt sind. Die Statistik erfasst in ihrer bekanntesten Ausprägung, wie viele Gegenspieler mit einer Fußballaktion (Pass oder Dribbling) überspielt wurden. Trotz der grundsätzlich simplen Idee gestaltet sich die Berechnung des jeweiligen Wertes komplex. Dabei werden wiederum Trackingdaten einbezogen: Der Abstand zum nächsten Gegenspieler, die Geschwindigkeit desselben sowie die Höhe des Passes spielen eine Rolle. Gleichzeitig wird auch die Anschlussaktion mit einberechnet (Impect, o. D.).

Packing



Abb. 8 Was ist Packing?
(Soccerdrills, 2016)

1. Evolution des Datenscouting und Abgrenzung der wesentlichen Kennzeichnungen

Der internationale Spitzenfußball unterliegt ständig einem technologischen Wandel. Durch die Weiterentwicklung von technisch-taktischen Strukturen, Veränderungen der ökonomischen und organisatorischen Rahmenbedingungen, aber auch durch technologische Innovationen entstehen neue Herausforderungen für Vereine und Verbände. Dies macht es erforderlich, vorhandenes Wissen über den Fußball immer wieder zu prüfen und neue Erkenntnisse zu gewinnen, die zum Erhalt und Ausbau der internationalen Konkurrenzfähigkeit beitragen. Dazu werden KPIs entwickelt, um beispielsweise einen Offensivverfolg einer Fußballmannschaft anhand von Pässen ins letzte Drittel zu bewerten. Die Analyse der technisch-taktischen Spielstrukturen ist ein wesentlicher Bestandteil des modernen Coachings im internationalen Profifußball (Carling et al., 2009).

Seit einigen Jahren stellen technologische Innovationen, insbesondere breite Verfügbarkeit von Positions- und Ereignisdaten sowie deren Verknüpfung in zentralen Datenbanken, neue Herausforderungen an die Art und Weise, wie diese Daten analysiert und interpretiert werden. Diese beinhalten beispielsweise Fragen, wie Vereine die Möglichkeiten nutzen können, um Spielverhalten zu untersuchen, Trainingsprozesse zu steuern und Transferentscheidungen zu verbessern. Eine gewichtige Komponente hierbei nimmt die Beschreibung der technisch-taktischen Aspekte des Spielgeschehens über Leistungsindikatoren ein (Hughes & Bartlett, 2002).

Lange wurden im Umfeld von Datendienstleistern und im Medienumfeld, aber auch der sportwissenschaftlichen Forschung vergleichsweise einfache Leistungsindikatoren wie Gesamtlaufrunden, Geschwindigkeitsprofile oder Heatmaps verwendet (Keyes & Nevill, 2014). Die wesentliche Herausforderung muss allerdings darin gesehen werden, aus den Rohdaten intelligente Leistungsindikatoren, KPIs, abzuleiten, die das Verständnis des Spiels erweitern, und substantielle Mehrwerte für die Spielanalyse und die Scoutingprozesse in Vereinen zu generieren. Die Aufgabenstellung besteht darin, die relevanten taktischen Strukturen im Fußball zu definieren, technisch verfügbar zu machen und die Qualität der Analyse nach wissenschaftlichen Kriterien abzusichern. Wesentlich ist, dass sich aus den entstehenden KPIs sportpraktisch relevante Aussagen für die Trainings- und Wettkampfsteuerung oder eine Spielerbewertung ableiten lassen (MacKenzie & Cushion, 2013; Weimar & Wicker, 2014). Im Rahmen der Auswertung und Interpretation von Daten haben sich im Laufe der Entwicklung von KPIs Bezeichnungen etabliert, die eine gezielte Beschreibung einer qualitativen Bewertung als Teil des Fußballspiels sind.

1. Evolution des Datenscouting und Abgrenzung der wesentlichen Kennzeichnungen

Expected Goals (xG) sind Statistiken, welche bestimmen wie groß bei einem bestimmten Schuss die Chance ist aus dieser Position ein Tor zu erzielen. Das xG-Modell wurde im Laufe seiner Entwicklung um Komponenten wie den Gegnerdruck erweitert. Dieser wird anhand der Positionierungen der ballnächsten Verteidiger in Bezug auf Ball, Raum, Zeit und Mitspieler mithilfe von Trackingdaten berechnet. Das Ergebnis ist eine xG-Grafik mit verschiedenen großen Kreisen und Punkten, welche die Güte einer Torchance darstellen, sowie ein daraus berechnetes, fiktives Endergebnis nach xG darstellen. Anhand des Expected-Goal-Modell wird also deutlich wie hoch eine Chance auf ein Tor wirklich war. Entscheidend ist dabei unter anderem, von wo auf dem Feld der Abschluss erfolgte, wie der Winkel zum Tor war und wie viele Gegenspieler sich noch zwischen Ball und Tor befanden. Der xG-Wert liegt immer zwischen 0 und 1. Mit ihm kann bei jeder Torchance klar bestimmt werden, wie hoch die Wahrscheinlichkeit war, dass der Ball von diesem Punkt aus im Tor landet. „xG 0,25“ bedeutet also: 25 von 100 Schüssen aus dieser Position führen zu einem Treffer. Die Werte beruhen auf großen Datenmengen und den dazugehörigen mathematischen Modellen (Bundesliga, 2021).

Expected Goals (xG)



Abb. 9 Darstellung der Expected Goals mit Punkten und Kreisen (Petty, 2019)

1. Evolution des Datenscouting und Abgrenzung der wesentlichen Kennzeichnungen

Aus der Ableitung zu den Expected Goals (xG) und Differenzierungen solcher Datensätze ergeben sich weitere KI-Ansätze, die in folgender Übersicht erläutert werden.

Erwartete Tore (xG) – die Anzahl der von einem Team oder einem Spieler zu erwartenden Tore, ausgehend von der Qualität und Quantität der Schüsse.

Erwartete Tore pro 90 (xG/90) – erwartete Tore pro 90 Minuten, die ein bestimmter Spieler spielt.

Nicht aus Elfm Metern erwartete Tore (npxG) – Gesamte erwartete Tore minus erwartete Tore aus Elfm Metern.

Erwartete Tore für (xGf) – die Anzahl der von einem Team erwarteten Tore, ausgehend von der Qualität und Quantität der Schüsse.

Erwartete Tore gegen (xGfa) – die Anzahl der von einem Team kassierten Tore, ausgehend von der Qualität und Quantität der Schüsse.

Erwartete Tore durch Zuspielen (xA) – die Anzahl der von einem Spieler erwarteten Torvorlagen, ausgehend von der Qualität und Quantität der Schüsse, die direkt aus seinem Pass hervorgehen.

Erwartete Punkte (xPts) – die Anzahl der Punkte, die ein Team voraussichtlich aufgrund der Daten der zu erwarteten Tore erzielt.

Expected Assists (xA) – die Anzahl der von einem Team oder einem Spieler zu erwartenden Assists, ausgehend von der Qualität und Quantität des vorletzten Passes.

Xhaka gegen Malli, Bundesliga-Saison 2014/15		
	Granit Xhaka	Yunus Malli
Schüsse pro 90	2,08	1,96
NpxG/90	0,06	0,21
XG/shot (aus dem Feld)	0,02	0,12
Tore pro 90	0,07	0,21

Tab. 4 X-Vergleich der Spieler Granit Xhaka (Borussia Mönchengladbach) vs Yunus Malli (FSV Mainz 05) aus der Spielzeit 2014/2015 (Petty, 2019)

1. Evolution des Datenscouting und Abgrenzung der wesentlichen Kennzeichnungen

Goals added (G+) ist innerhalb der Erfassung rund um die qualitative Datenanalyse ein recht neuer Wert. G+ ist eine Weiterentwicklung der Expected Goals (xG) mit dem Unterschied, dass bei dieser Funktion nicht nur die Torabschlüsse, sondern jede einzelne Ballaktion zu einer Tor- und Gegentorwahrscheinlichkeit zugeordnet wird. Unterschiedliche Anbieter versehen den Wert mit unterschiedlichen Bezeichnungen und Definitionen. Das Ziel dieser Herangehensweise ist es aber immer, die Torwahrscheinlichkeit zu verifizieren, die durch eine Aktion (z. B. gewonnener/verlorener Zweikampf) verändert wird. Erhöht sich die eigene und/oder verringert sich die gegnerische Torwahrscheinlichkeit, wird das dem entsprechenden Spieler, der die Aktion ausführte, gutgeschrieben. Dadurch erfährt jeder Schuss, jeder Pass, jeder Zweikampf oder Ballgewinn/-verlust eine Gewichtung und Bewertung, die letztendlich die Bewertung aller Aktionen jedes einzelnen Spielers ermöglicht. Hierzu zwei Übersichten der US-amerikanischen Plattform American Soccer Analysis, die zum einen die besten G+-Werte der MLS und zum anderen die Entstehung dieser Erhebung visualisiert.

Goals added (G+)

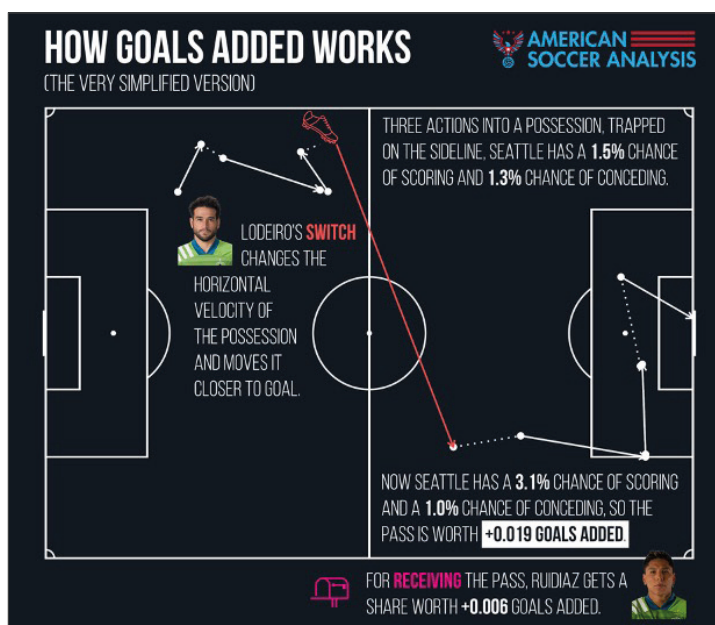


Abb. 10 American Soccer Analysis – Beschreibung Goals Added+
(American Soccer Analysis, o. D.)

1. Evolution des Datenscouting und Abgrenzung der wesentlichen Kennzeichnungen



Abb. 11 Goals Added+ – best players 2019
(American Soccer Analysis, o. D.)