

# Methoden des Machine Learning am Beispiel von Wettmärkten

## Masterarbeit

zur Erlangung des akademischen Grades „Master of Science (M. Sc.)“ im  
Studiengang Wirtschaftswissenschaft der Wirtschaftswissenschaftlichen Fakultät  
der Leibniz Universität Hannover

vorgelegt von

Name: Ahadi



Vorname: Sadaf



Prüfer: Prof. Dr. H.-J. von Mettenheim

Hannover, 26.02.2018

# Inhaltsverzeichnis

---

<b>Abstract</b>	
<b>Abbildungsverzeichnis</b> .....	<b>V</b>
<b>Tabellenverzeichnis</b> .....	<b>VI</b>
<b>Abkürzungsverzeichnis</b> .....	<b>VII</b>
<b>1 Einleitung</b> .....	<b>1</b>
<b>2 Theoretische Grundlagen</b> .....	<b>4</b>
2.1 Machine Learning .....	4
2.2 Fuzzy Set Theory .....	7
2.3 Wettmärkte .....	10
2.3.1 Entwicklung von Sportwetten .....	11
2.3.2 Fußball und Fußballwetten.....	13
<b>3 Herleitung der Forschungslücke</b> .....	<b>17</b>
<b>4 Methodik</b> .....	<b>20</b>
4.1 Fuzzy Rules und Fuzzy Inference Systems.....	20
4.2 Adaptive Network-based Fuzzy Inference System.....	23
4.3 Conditional Fuzzy.....	27
4.4 Relevance Vector Machines .....	33
<b>5 Ergebnisse</b> .....	<b>38</b>
5.1 Dataset.....	38
5.2 Modell-Synthese .....	40
5.3 Diskussion.....	41
<b>6 Fazit und Ausblick</b> .....	<b>60</b>
<b>Literaturverzeichnis</b> .....	<b>61</b>
Ehrenwörtliche Erklärung.....	69

# 1 Einleitung

In der heutigen Zeit, in der technische Entwicklung immer weiter und immer schneller voranschreitet, nehmen Computersysteme im Alltag der Menschen stetig an Bedeutung zu. Unabhängig von dem Standort, in dem man sich befindet, ist man von Maschinen umgeben, die mit Hilfe der sogenannten *künstlichen Intelligenz* betrieben werden. Angefangen von Zuhause, gibt es den Staubsaugerroboter, der bei den Tätigkeiten im Haushalt eine große Bereicherung ist. Dieser ist ein vollautomatischer und selbstständiger Staubsauger, der über eine Hindernis- und Treppenerkennung sowie über eine Sprachsteuerung verfügt.<sup>1</sup> Unterwegs gibt es heutzutage autonome Metros, die im Jahr 2015 ca. eine Milliarde Fahrgäste im europäischen Nahverkehr befördert haben. Diese Bahnen verfügen über viele Vorteile, weswegen sie bei den Fahrgästen sehr beliebt sind: kürzere Taktung und damit Zuwächse bei der Kapazität, höhere Energieeffizienz, bessere Verfügbarkeit, gute Pünktlichkeitswerte und hohes Sicherheitsniveau.<sup>2</sup> Doch auch auf der unscheinbaren Ebene spielt künstliche Intelligenz eine wichtige Rolle, insbesondere für Unternehmen. Anhand von ubiquitären Computersystemen werden jegliche Informationen über getroffene Entscheidungen, wie z. B. Einkäufe im Supermarkt, oder Suchvorgänge im Internet gespeichert. Mit Hilfe der künstlichen Intelligenz wird diese Flut an Informationen nach Relevanz gefiltert, gespeichert und weiterverarbeitet (Witten & Frank, 2005, S. 4). Diese nützlichen Informationen werden von den Unternehmen genutzt, um die Bedürfnisse ihrer Kunden<sup>3</sup> besser einschätzen, ihnen auf diese Bedürfnisse angepasste Produkte anbieten und somit ihre Umsätze steigern zu können.

Nicht nur Unternehmen nutzen die Vorteile der künstlichen Intelligenz, sondern diese werden auch im Bereich der Wettmärkte zur Vorhersage von Wetterfolgen genutzt. Besonders Prognosen von Sportwetten wurden in der Literatur ausführlich untersucht,

---

<sup>1</sup> Saugroboter Kobold VR200, <https://kobold.vorwerk.de/saugroboter>, aufgerufen am 10.12.2017

<sup>2</sup> Allianz pro Schiene, 2016, <https://www.allianz-pro-schiene.de/presse/pressemitteilungen/uebersicht-selbstfahrende-metros-europa>, aufgerufen am 10.12.2017

<sup>3</sup> Zur Vereinfachung der Darstellung und aufgrund der besseren Lesbarkeit wird auf die weibliche Form von Bezeichnungen etc. verzichtet. Die maskuline Bezeichnung ist daher immer auch für die weiblichen Personen gültig zu sehen.

denn sie sind wichtig für Sportfans, Teammanager, Sponsoren, Medien sowie für die zunehmende Anzahl an Wettteilnehmern (Vlastakis, Dotsis & Markellos, 2009). Da diese Art von Prognosen von externen Ereignissen beeinflusst werden, nutzen Forscher auf diesem Gebiet komplizierte Algorithmen, wie z. B. die Bayes'sche Statistik oder Methoden der künstlichen Intelligenz. Allerdings sind diese sehr komplex und erschweren die Ableitung einer Prognose für viele Anwender, die genau wissen müssen, wie die abhängige und die unabhängige Variablen miteinander interagieren. Mit Hilfe der sogenannten *Fuzzy-Regeln*<sup>4</sup> kann diese Komplexität beseitigt werden (Serpinis, Hassanniakalager, Stasinakis & Verousis, 2017).

Serpinis et al. (2017) haben in einer Studie die Conditional Fuzzy Inference (CF) mit der Relevance Vector Machines (RVM), welche eine Bayes'sche Technik ist, verknüpft, um die Vorteile der CF zu untersuchen. Dieses Modell wurde für die Vorhersage der Ausgänge sowie der Anzahl der Tore bei den Fußballspielen der drei größten Fußballmannschaften - der englischen Premier League, der italienischen Serie A und der spanischen La Liga - von 2005 bis 2015 angewandt. Die Studie hat ergeben, dass die CF-RVM-Kombination genauer ist als die RVM. Des Weiteren scheint CF eine robuste Methode zu sein und erhöht zudem die Gewinnchance bei Wettspielen.

Die vorliegende Arbeit knüpft an die Studie von Serpinis et al. an und erweitert sie, indem die Ausgänge sowie die Anzahl der Tore der deutschen Bundesliga von 2005 bis 2017 mit demselben Modell prognostiziert werden.

Die Masterarbeit ist folgendermaßen aufgebaut: Im nachfolgenden zweiten Teil werden zentrale Begrifflichkeiten definiert und erörtert. Im Anschluss daran wird in Kapitel drei die Forschungslücke hergeleitet. Kapitel vier beschäftigt sich mit der Methodik der Studie. Dabei wird auf *Fuzzy Rules (FR)* und *Fuzzy Inference Systems (FIS)*, *Adaptiv Network-based Fuzzy Inference System (ANFIS)*, *Conditional Fuzzy (CF)* und *Relevance Vector Machines (RVM)* eingegangen. Im fünften Teil werden die Ergebnisse der Studie präsentiert und interpretiert. Kapitel sechs schließt die Arbeit mit einem Fazit und Ausblick ab.

---

<sup>4</sup> Die Fuzzy-Regeln werden in Kapitel 4 näher erläutert.

Abbildung 1 gibt einen grafischen Überblick über das methodische Vorgehen der vorliegenden Arbeit.

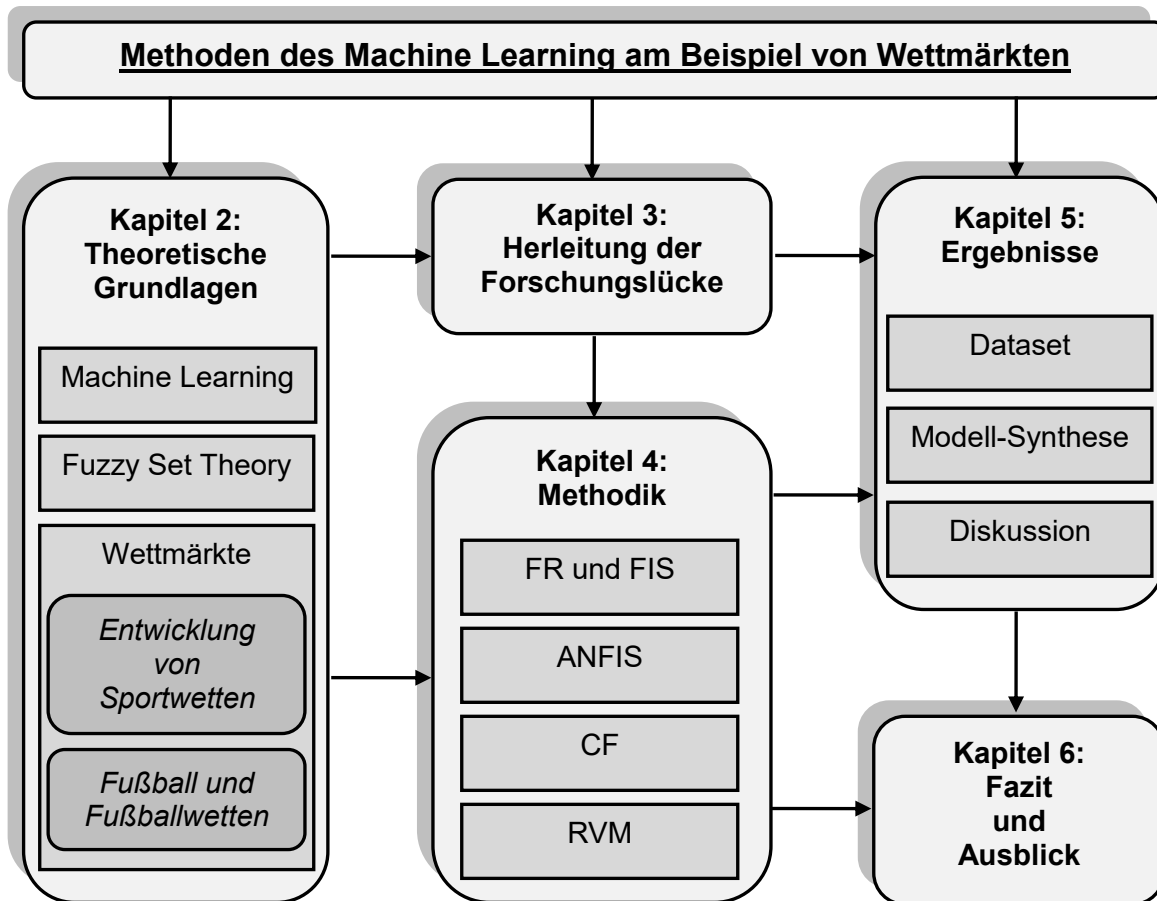


Abb. 1: Überblick des methodischen Vorgehens (Eigene Darstellung)

## 6 Fazit und Ausblick

Diese Arbeit knüpfte an die Studie von Sermpinis et al. (2017) an, die das Konzept von CF einführten und ihren Nutzen demonstrierten. CF erzeugt eine Reihe von FRs in dem In-Sample und schätzt ihre durchschnittliche Firing Strength. Diese Regeln werden in einer Rangfolge geordnet und in dem Out-of-Sample angewendet. CF generiert Vorhersagesignale an Punkten, an denen starke Regeln in der Nähe sind und einen endogenen und einen exogenen Schwellenwert erfüllen. Das Prognosesignal ist ein gewichteter Durchschnitt der stärksten Regeln (Sermpinis et al., 2017).

Sermpinis et al. (2017) haben CF empirisch getestet, indem sie sie mit RVM kombiniert haben und sie für die Prognose der Spielergebnisse, der Spielergebnisse durch das Asian Handicap und der Anzahl der Tore pro Spiel der Premier League, La Liga und Seria A angewandt haben. Ihre Leistung wurde mit denen einer RVM und einer RVM kombiniert mit ANFIS, welches der bekannteste Unconditional-Fuzzy-Ansatz ist, verglichen. Die Ergebnisse der Studie haben gezeigt, dass die CF-RVM-Kombination die Genauigkeit von RVM verbessert und das RVM-ANFIS-Modell übertrifft. Des Weiteren konnte festgestellt werden, dass CF höhere Renditen im Glücksspiel bietet als die anderen Modelle. Alle CF-Ergebnisse erschienen robust für zwei Fußball-Saisons (Sermpinis et al., 2017).

Wie die Studie demonstriert, bringt CF viele Vorteile mit sich. Sie kann ein nützliches Werkzeug für Anwender und Akademiker sein, die sich mit Problemen befassen, die komplexe nichtlineare Techniken erfordern. In Bereichen wie Medizin, Finanzen und Wirtschaft kann sie das Verständnis der Forscher für die zugrundeliegende Beschaffenheit der Serien verbessern. In Bereichen wie der Meteorologie oder dem Fußball-Glücksspiel kann sie Entscheidungsregeln bieten, die einfach und interpretierbar sind. Die CF bietet eine attraktive Alternative zu den zahlreichen Unconditional-Fuzzy-Inferenz-Ansätzen, die in der Literatur dominieren, die den Datensatz erfassen, jedoch zu einer Verschlechterung der Leistung des Systems führen (Sermpinis et al., 2017).