Bergische Universität Wuppertal

Fakultät für Maschinenbau und Sicherheitstechnik

Masterthesis

THEMA: STATISTISCHE DATENANALYSE, MODELLIERUNG UND SIMULATION VON SPURWECHSELMANÖVERN AUF AUTOBAHNEN

im Studiengang Sicherheitstechnik
beim Fachgebiet Verkehrssicherheit und Zuverlässigkeit

zur Erlangung des akademischen Grades
Master of Science

Autor: Kokou Mawuena Apedoh
MatNr. 1613786

Bearbeitungszeitraum: 16.11.2020 bis 15.04.2021

1. Betreuer: Jun.-Prof. Antoine Tordeux
2. Betreuer: Basma Khelfa

Abstract

This Thesis contributes to predict a lane-changing maneuver on Highways truth a real HighD Dataset recorded in Germany with a drone. Therefore we proceed analysing the raw data using rigorous multivariate Statistical methods to visualize and to understand the relations between the variables and to distinguish, wich kind of the Variables played some important role according to the lane changing. The major Goals remain to improve the Quality and the safety of the circulation truth reducing the accidents due to the lane changing manipulation on the Highways.

Abstract

Diese Arbeit trägt dazu bei, ein Spurwechselmanöver auf Autobahnen vorherzusagen. Der in dieser Arbeit verwendete Datensätze stammt aus der HighD Datenbank und Sie wurden mit einer Drohne erfasst. Es geht in erster Linien um die Explorative Multivariable Datenanalyse mit dem Ziel die Datensätze besser zu verstehen, also die Beziehungen zwischen den Variablen einerseits zu verstehen und zu visualisieren und andererseits erkennen, welche Variablen eine entscheidende Rolle spielen, wenn es um den Spurwechsel geht. Darüberhinaus wurde anhand Entscheidungsbaum Methoden und logistischer Regression die zweiten Teil der Arbeit durchgeführt. Die Hauptziele bleiben die Verbesserung der Qualität und die Sicherheit des Verkehrsflußes, um die Unfälle aufgrund der Spurwechselmanöver zu verringern.

Inhaltsverzeichnis

Vorwort ii

Inhaltsverzeichnis vii

1 Einleitung 1

2 Einführung in das autonome Fahren 3
2.1 Modelle zur Beschreibung des Fahrverhaltens . . . 3
2.1.1 Drei-Ebenen-Modell für zielgerichtete Tätigkeiten des Menschen (Rasmussen, 1983) . . . 3
2.1.2 3-Ebenen-Modell der Fahrzeugführungsaufgabe (Donges, 1994) . . . 4
2.2 Fahrerassistenz auf Autobahnen . . . 6
2.2.1 Sichtverbesserungssysteme . . . 6
2.2.2 Frontalkollisionsschutzsysteme . . . 11
2.2.3 Lane Departure Warning . . . 13
2.2.4 Lane Departure Warning-System(LDW) . . . 13
2.2.5 Advanced Lane Departure WarningSystem (ALDW) . . . 14
2.2.6 Lane Keeping Support . . . 14
2.2.7 Lane Departure Prevention (LDP) . . . 14
2.3 Vorhersagende Spurwechselnde Modelle “Lane changing Model“ . . . 14
2.3.1 Struktureller Aufbau von dem Lane changing Model . . . 17
2.3.2 Spur Auswahl Modelle . . . 18
2.3.3 gap acceptance model von Ahmed . . . 19
2.4 Relevante Methoden . . . 20
2.4.1 Angewandte Lineare Regression . . . 20
2.4.2 Regression Logistik . . . 21
2.4.3 Prinzipal Komponente Analysis . . . 23
2.4.4 Entscheidungsbäume . . . 23

3 Statistische Analyse der HighD-Messdaten 25
3.1 Beschreibung der Messdaten . . . 25
3.1.1 explorative Datenanalyse . . . 26
3.1.2 Korrelationsanalyse . . . 28
3.1.3 Hauptkomponente Analyse der Messedaten . . . 28
3.1.4 Zusammenfassung der Beschreibende Statistik . . . 33
3.2 Decision tree Analyse zur Auswahl der erklärenden Variablen . . . 34
3.3 Vorhersagen von Spurwechsel Anhand Logistische Regression . . . 35
3.3.1 Train-, Test-, Validation Datensätze . . . 35
3.3.2 Ergebnis . . . 35
3.3.3 Evaluation . . . 36
3.4 Netlogo . . . 37
3.4.1 Oberfläche und Simulationsparameter . . . 38
3.4.2 Durchführung . . . 39

3.4.3 Ergebnis und Interpretation 40

4 Schluss **41**

Literaturverzeichnis **43**

Abkürzungsverzeichnis **47**

Bilderverzeichnis **49**

Tabellenverzeichnis **51**

Eidesstattliche Erklärung **53**

Anhang **55**

1 Einleitung

In Deutschland laut des Statistischen Bundesamtes (Destatis), kommt alle 9 Stunden, ein Mensch durch erhöhte Geschwindigkeit ums Leben. Im Jahr 2019, laut Polizeiermittlungen, wurde festgestellt, dass 32 Prozent der Menschen, die bei einem Verkehrsunfall ums Leben kamen, in einem Unfall verwickelt waren, wobei der Fahrer nicht auf die Geschwindigkeit geachtet hatte. Es gab in den vergangen Jahren insgesamt 963 Tote, 53 687 Verletzte und von diesen 13769 schwere Verletzte. [1] Ein Zehntel dieser Unfälle passiert durch das Wechseln der Spur oder beim Auffahren auf die Autobahn. Dies führt zur Verkehrsverzögerung oder Verkehrsstau und hat sowohl eine negative Wirkung auf die Ankunftszeit als auch auf die Wirtschaft. [2] Es zeigt sich, dass menschliche Fehler wie zum Beispiel Ermüdung, Unaufmerksamkeit des Fahrers und in vielen Fällen, das Einschlafen des Fahrers während der Fahrt, einen großen Teil der Autounfälle ausmacht. Studien haben gezeigt, dass der Abstand zwischen den Fahrzeugen sowohl für die Sicherheit als auch für die Autobahnkapazität von größter Bedeutung sind, um eine Kollision zu vermeiden. Um sichere Fahrzeugmanövern zu gewährleisten, sollte der Abstand zwischen den Fahrzeugen groß genug sein. [3]
Um diese Probleme zu lösen, die Sicherheit im Straßenverkehr zu sichern und die Anzahl der Unfälle zu verringern, wurden Bildverarbeitungssysteme entwickelt. Solchen Systeme bestehen aus einer installierten Videokamera oder Sensoren im Innenraum des Fahrzeugs, zur Erfassung der Umgebung mit dem Ziel, nützliche Informationen für den Fahrer bereitzustellen, Beispielsweise Geschwindigkeitsbegrenzung, Ampeln, Hindernisse. Auch wurden Warnsysteme entwickelt. [4] [5]
Es ist den Automobilhersteller ganz bewusst und Sie haben Sorge getragen die Prädiktive Fahrerassistenzsysteme und Sicherheitssysteme in das Auto zu bauen um die Sicherheit beim Fahren zu gewährleisten. Der Volkswagen Konzern verfügt in diesem Zusammenhang beispielsweise über automatische Distanzregelung ACC, die Multicolusionbrems, „Side Assist“, „Front Assist“ mit City-Notbremsfunktion, Spurhalteassistenz „lane Assist“ [6]
Der ADAC Konzern ist der Meinung, dass die elektronischen Fahrassistenzsysteme von großer Bedeutung sind, um die Anzahl der Unfälle weiter zu mindern oder deren Schaden gering zu halten. Nach Ansicht des ADAC ist das Fahrerassistenzsysteme wichtig. Es soll dabei den Fahrer bei der Verkehrssicherheit unterstützen und den Fahrkomfort gewährleisten. Dadurch wir ein guter Verkehrsfluss ermöglicht. [7]
Die Auswertung des Prädiktive Fahrerassistenz- und des Sicherheitssysteme haben sich als positiv im realen Leben erwiesen. So konnten ACC-Systemen bis zu 7 Prozent der Unfälle mit Personenschaden und schwerem Sachschaden vermeiden. Besonders das hohe System Lane Departure Warning (LDW) ist als Komfortsystem für die Autobahn gedacht. Mittels einerVideokamera wird die Fahrspur erfasst und ausgewertet. Der Fahrer wird bei unbeabsichtigtem

Befahren der Fahrbahnmarkierung mit einem akustischen Warnsignal darauf aufmerksam gemacht, eine Kurskorrektur vorzunehmen. Die AZT-Studie gemeinsam mit MAN im Rahmen von Safe Truck ergab, dass 8 % der Lkw-Unfälle mit Personenschaden oder schwerem Sachschaden LGS-relevant sind. Rund die Hälfte davon bzw. 4 % dieser Unfälle könnten mit dem heutigen LGS vermieden werden. [7]

Es gibt viele Gründe weshalb eine Überholung vorgenommen wird, entweder aufgrund des Geschwindigkeitsunterschied, Abstand zwischen dem folgenden Fahrzeug oder selbst Beschleunigung, sowie das Auffahren von anderen Fahrzeugen. Viele Parametern spielen also für einen Spurwechseln eine entscheidende Rolle und manche eher weniger. Insgesamt geht es um 25 variablen die eine Rolle spielen.
Hauptziele der Arbeit ist es, prädiktive Modelle zu entwickeln, um vorhersagen zu können, wann ein Spurwechsel stattfinden sollte. Dafür wird anhand multivariable Datenanalysen, Methoden untersucht, wie die Variablen miteinander zusammen hängen, wenn es um den Spurwechseln geht. Ferner werden verschiedene Methoden genutzt, um herauszufinden welche Parameter am meisten eine Rolle spielen. Darüber hinaus werden verschiedene Algorithmen verwendet, um vorhersagende Modell zu entwickeln.

2 Einführung in das autonome Fahren

2.1 Modelle zur Beschreibung des Fahrverhaltens

Die Führung eines Fahrzeugs besteht aus einer komplexen und verantwortungsvollen Überwachungsaufgaben und Regelungsaufgaben. Um dem Fahrer eine angenehme Arbeitsbedingung zu verschaffen, muss die Auslegung der Komponenten an die Leistungsfähigkeit des Menschen angepasst werden. In der zweiten Hälfte des 20. Jahrhunderts begann man, [8] Erkenntnisse über das Verhalten von Fahrern, während der Fahraufgabe zu sammeln und in Form von Fahrermodellen zusammenzufassen. [8]

2.1.1 Drei-Ebenen-Modell für zielgerichtete Tätigkeiten des Menschen (Rasmussen, 1983)

Das Modell wurde im Jahr 1983 von Rasmussen entwickelt, stammt aus der Ingenieurpsychologie und erfasst „von alltäglichen Routinesituationen über unerwartete Herausforderungen bis hin zu seltenen kritischen Störfällen “, und stellt die unterschiedlichen Phasen des menschlichen Lernverhaltens dar. Diese Drei-Ebenen-Struktur ist in Bild 2–1 links dargestellt.

Dies Modell dient zur Darstellung des mentalen und kognitiven Prozesses. Die Ebene des wissensbasierten Verhaltens (knowledge-based behaviour), erfasst die menschliche Kognition und darunter verstehe man die Information Assimilation in Form von Wissen und Regeln, die auch Leistungsfähigkeit Ebene genannt wird; wie zum Beispiel Fahrerfahrung. Das heißt auf der Basis des erworbenen Wissens, in einer bestimmten Situation, verschiedene Handlungsmöglichkeit im Kopf abzuspielen, um die richtige Entscheidung zu treffen. [8] [9]

das „regelbasierten Verhaltens “(rule-based behaviour) ist damit gekennzeichnet, dass die Situation schon häufiger aufgetreten ist, sodass eine effektive Handlungsweise entwickelt wurde. [8]

Die Erfahrung und die Routine Ausübung einer Tätigkeit bringt Menschen auf die Ebene, des Fertigkeitsbasierten Verhaltens, die nicht anders ist als Automatisierung oder reflexartige Ebene, die sogenannten Sensomotorische Fähigkeiten. Die Erreichung dieser Ebene, hat wiederum einen eindeutigen positiven Einfluss auf die Leistungsfähigkeit. Darüber hinaus, erhöht sich die unbewusste Kontrolle bei der Ausführung einer Tätigkeit wie zum Beispiel, das Fahrradfahren. [9]

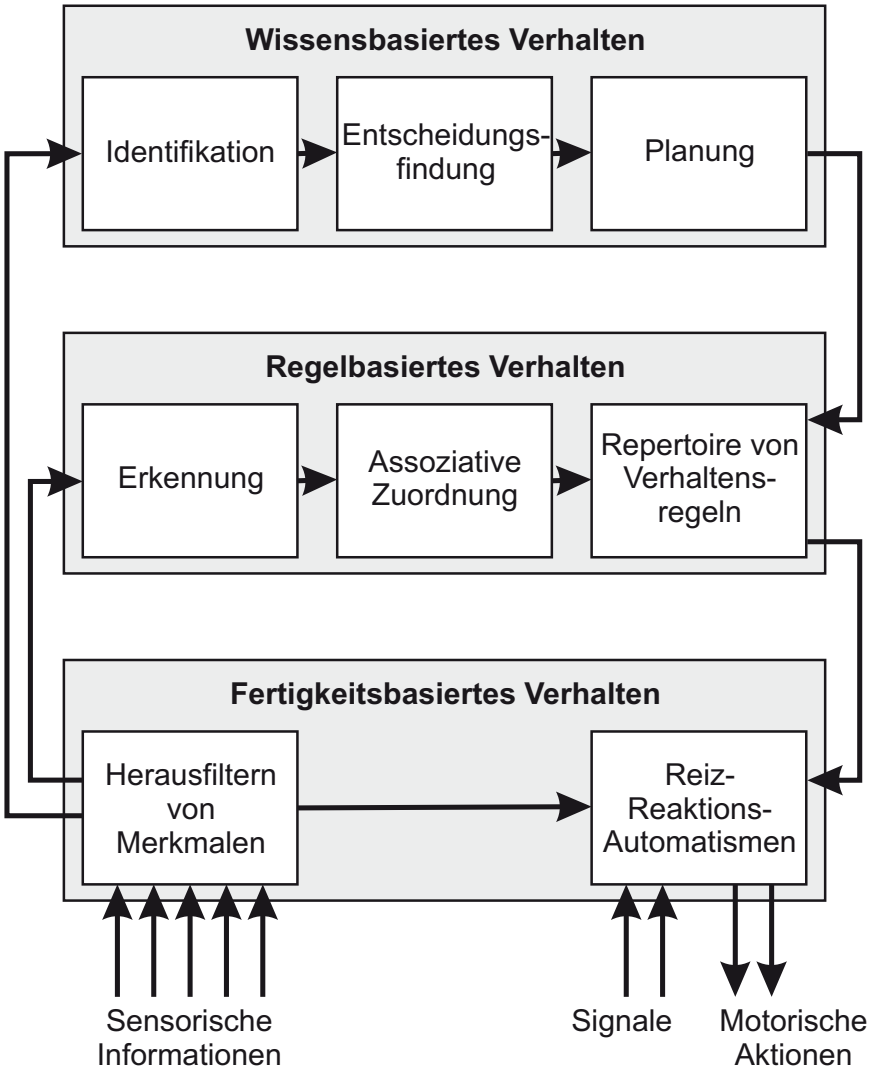


Bild 2.1: Drei-Ebenen-Modell nach Rasmussen [8]

2.1.2 3-Ebenen-Modell der Fahrzeugführungsaufgabe (Donges, 1994)

Donges entwickelt im 1982 die Drei-Ebenen-Hierarchie der Fahraufgabe. Er unterscheidet für die Führung eines Fahrzeuges 3 Ebenen, und zwar Navigation, Führung und Stabilisierung. [10]

Die Ebene der Navigation besteht aus der Auswahl einer geeigneten Fahrstrecke und der Abschätzung des voraussichtlichen Zeitbedarfs. Im Falle eines Störeinflusses, wie zum Beispiel Unfälle, Baustellen oder Verkehrsstauungen, soll eine Routenplanänderung vorgenommen werden. Die Navigationsebene ist direkt abhängig von den veränderte Umgebungsbedingungen. Die Navigations-Ebene entspricht der wissensbasierten Ebene nach Rasmussen. Donges unterschei-

det zwei Prozesse, den diskreten und den dynamische Prozess. DieNavigationstätigkeit gehört zu den diskreten Prozessen; da es Zeitlich diskret ist. Er Ordnet die Ebene der Führung und Stabilisierung als dynamischer Prozesse zu. [10]. [11]
Die Führungsebene hängt sowohl von den Fähigkeiten und Fertigkeiten als auch von der Verfassung des Fahrers ab, genau gesagt von allen drei Ebenen der Rasmussen ab. Die Stabilisierungsebene ist eine Kompensation Tätigkeiten der Fahrer zwischen den Führungsgrößen und den Ist-Größen, die sich Kontinuierlich durch Eigenbewegung sowie bewegte fremd Objekte auf der Strecke ändern. [10]. [11]

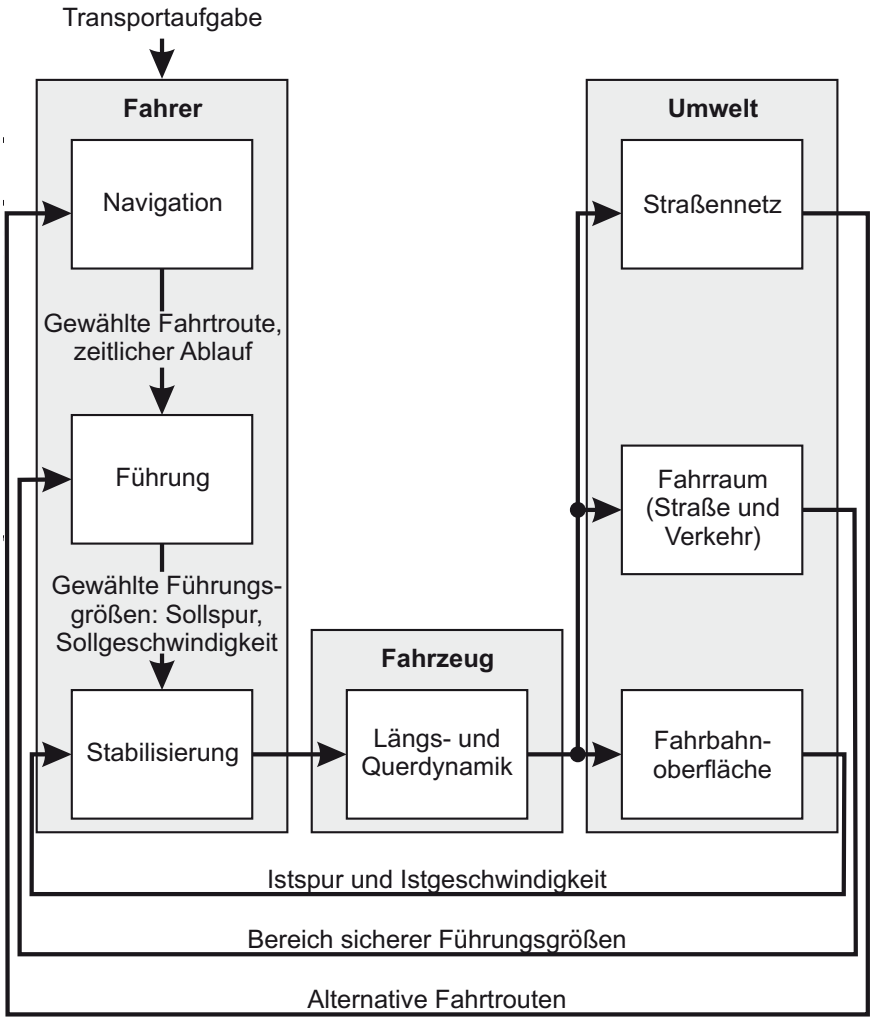


Bild 2.2: Drei-Ebenen-Modell nach Donges [8]

2.2 Fahrerassistenz auf Autobahnen

Der Begriff der Fahrerassistenz wurde zum ersten Mal durch das EUREKA Forschungsprojekt PROMETHEUS verwendet. [12] Darunter versteht man Fahrzeugsysteme, die den Fahrer sowohl für das Manövern, Fahrzeugstabilisierung als auch für die Führung des Fahrzeuges unterstützen, und zwar durch Information, Warnung und Sicherstellung der Sicherheit. [13] [14] Sinnvoller wäre es den Fahrer zu unterstützen, genau dort wo seine Fähigkeiten am meisten nachlässt oder begrenzt ist, wie zum Beispiel beim Schätzen der Geschwindigkeit oder Entfernung in solchen Situationen mit schlechter Sicht, Abstand und die Spurlage. Unter Fahrerassistenz versteht man unter anderem Querführungssysteme wie Heading Control, Geschwindigkeitsregler wie Cruise Control und Stabilisierungssysteme wie ABS (Anti Blocking System), ASC (Anti-Skid-Control) oder DSC (Dynamic-Stability-Control). [13]

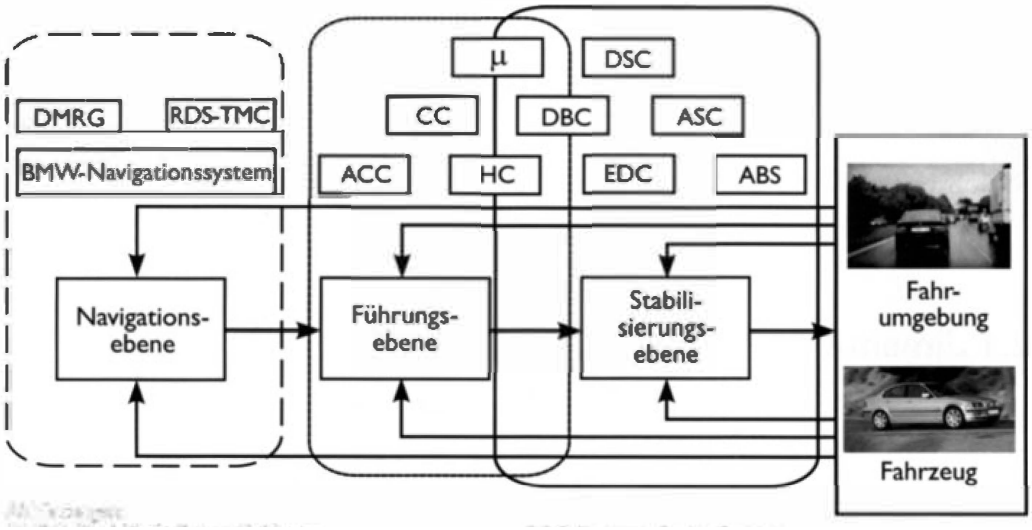


Bild 2.3: Fahrassistenzsysteme im Zusammenhang mit Fahr-aufgaben [13]

2.2.1 Sichtverbesserungssysteme

Eine Analyse des Institutes für Fahrzeugsicherheit in München, aus dem Jahr 1997 hat ergeben, dass knapp 14596 der 43789 Unfälle, im Jahr 1995 bei denen Fußgänger verletzt wurden, im dunklen stattgefunden haben. 802 der 1336 Unfälle endeten tödlich. [15] Die Bundesanstalt für Straßenwesen im Jahr 2005, hat die Daten der Nachtunfälle näher betrachtet und hat die Daten nach folgende Kriterien sortiert:

- Sortierung nach Bundesländern und nach Ortslagen (Innerorts, Bundesstraßen, Autobahnen),
- Verkehrsunfälle nach der Zeit

- Unfalltyp und Unfallart sowie Unfallumstände,
- Unfall teilnehmern Eigenschaften (Alter, Geschlecht)
- „Hauptverursacher der Unfälle nach Art des Verkehrsteilnehmers (Fußgänger, Pkw, Fahrrad, Moped/Mofa . . .)“.

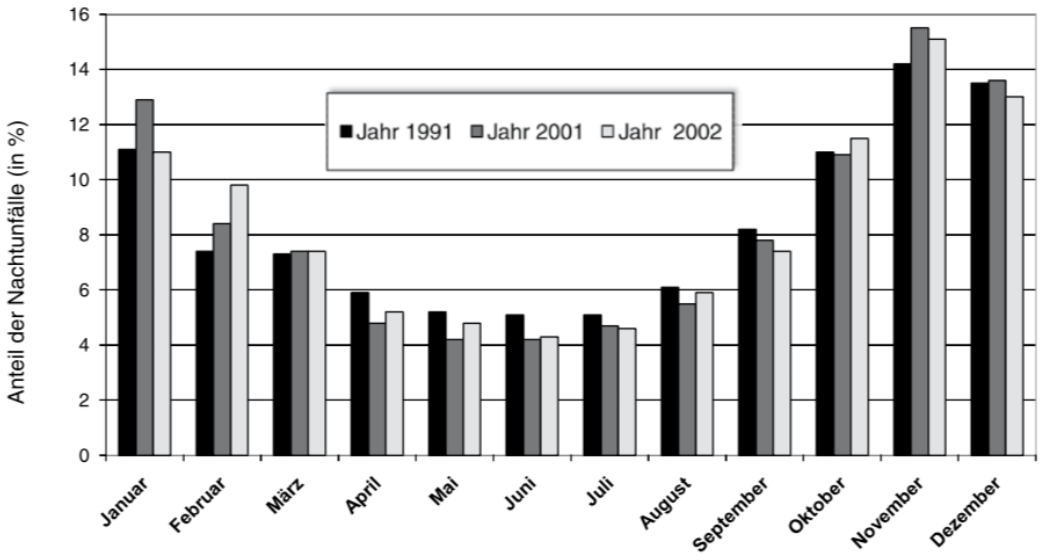


Bild 2.4: Häufigkeitsverteilung von Nachtunfall [16]

In Abbildung. 2.4 werden die Anteile der Nachtunfälle über die Monate im Jahresverlauf von 1991 bis 2002 veranschaulicht. Die Verteilung nach Monaten zeigt, dass der höchste Anteil in den Wintermonaten von Oktober bis Februar zu sehen ist und in den Sommermonaten von Mai bis Juli minimal ist. Das Verhältnis zwischen beide Bildern ergibt, dass der Anteil der Nachtunfälle in den Monaten von November bis Januar dreimal so viel ist, wie in den Monaten Mai bis Juli. Dies hat folgende Gründe: [17]

- Verschlechterung der Sichtbedingungen im Dunklen
- die allgemeine Wetterbedingung und Fahrbahneigenschaften im Winter

Analysiert man die Ursachen für Nachtunfälle im Jahr 2002 genauerer, zeigt in Abbildung 2.5, dass der Anteil der Nachtunfälle bei Schnee, Eis und Regen an der gesamten Unfallzahlen 27 Prozent beträgt.
Der Nachtanteil der jeweiligen Unfallursache ist bemerkenswert hoch bei Schnee und Eissowie bei Nebel und Wild auf der Fahrbahn mehr als 55 Prozent. Die Unfallursachen deutet darauf hin, dass Objekte auf der Fahrbahn im dunklen zu spät erkannt werden. [17]

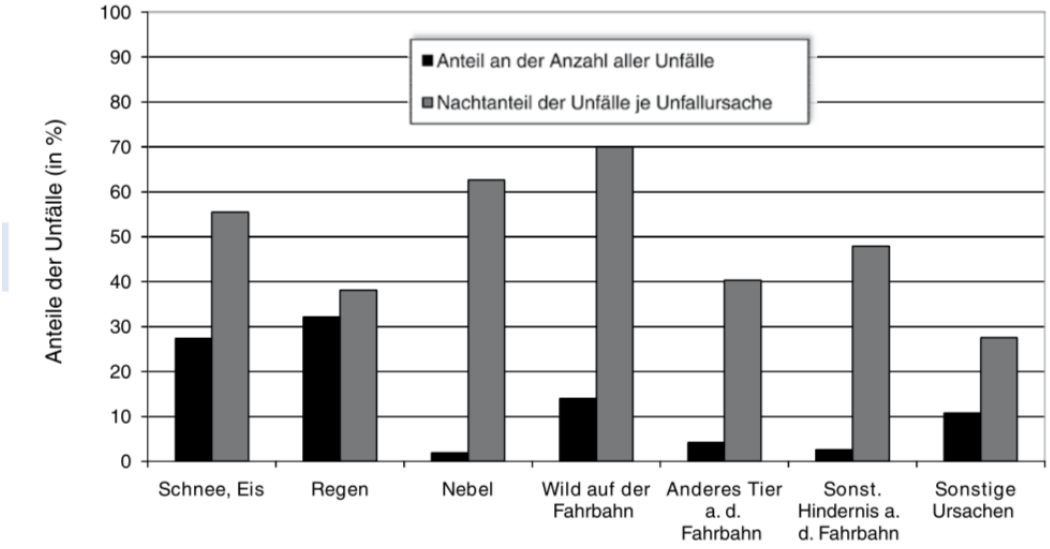


Bild 2.5: Nachtunfall Ursachen [16]

Lichtquellenentwicklung

Die Scheinwerfersysteme verwenden als Lichtquellen entweder Lampen aus Halogene, aus Xenon oder aus LED. Die im Tabelle 2.1 und 2.2 veranschauliche Lichtquellen Eigenschaften, deutet darauf hin, dass heutzutage viel mehr LED Lampen verwendet werden, weil es energieeffizienter, umweltschonender, günstiger und moderner ist, als die Scheinwerfer aus Halogenglühlampen. Die LED-Komponenten haben öfter eine längere Lebensdauer als das Kraftfahrzeug selbst. „Die liegt in der Größenordnung von 8000 bis 10.000 Stunden “ [17]. Der Ausfall sowie das häufige Wechseln der Lichtquellen können somit weitgehend ausgeschlossen werden. Die LED- Bauelemente als Halbleiter Lichtquellen, können sehr schnell und jederzeit gedimmt und ausgeschaltet werden. Diese ermöglichen, dass die Scheinwerfer ganz schnell, und zwar in einem Millisekunden-Takt das Licht verteilen und auf die Fahrbahn anpassen. Dank seiner Bauformen, können die LED-Bauelemente verschiedene Lichtfunktionen wie zum Beispiele Fernlicht, Abblendlicht, Kurvenlicht usw.in einer kompakten Baugruppe gebaut werden. [17]

Tabelle 2.1: Eigenschaften der Lichtquelle eines Frontscheinwerfers

Lampentyp	Lichtstrom	Max. Leuchtdichte	Lichtausbeute	Farbtemperatur
Halogenglühlampe (H7)	1500 lm	30 Mcd/m²	25 lm/W	3200 K
Xenonlampe (D2S)	3200 lm	90 Mcd/m²	90 lm/W	4200 K.
LED (kaltweiß)	150–1500 lm	20 Mcd/m²	65 lm/W	4000 bis 6000 K

Tabelle 2.2: Sichtbarkeitsweite Lichtquelle eines Frontscheinwerfers

Lampe mit	Sichtbarkeitsweite nach	Sichtbarkeitsweite unter 0°nach	Sichtbarkeitsweite unter 20°nach
Halogenglühlampe (H7)	70 m	63 m	18,3 m
Xenonlampe (D2S)	85 m	80 m	25,8 m.

Adaptive Lichtverteilung

Ein Zusammenhang zwischen den Bremsweg und der Fahrgeschwindigkeit in Bild 2.6 zeigt uns, dass eine Sichtbarkeitsweite von 85m, was als Maximum angesehen wird, mit Scheinwerfern aus Xenon Lichtquelle, eine maximale Fahrgeschwindigkeit von etwa 90 km/h in der Nacht sicherstellt. Moderne Scheinwerfer aus Halogen mit einer Sichtbarkeitsweite um 65 m entsprechen eine Fahrgeschwindigkeit, von etwa 75 km/h. Eine Sichtverbesserung auf Basis der Lichtquellen reicht nicht aus.

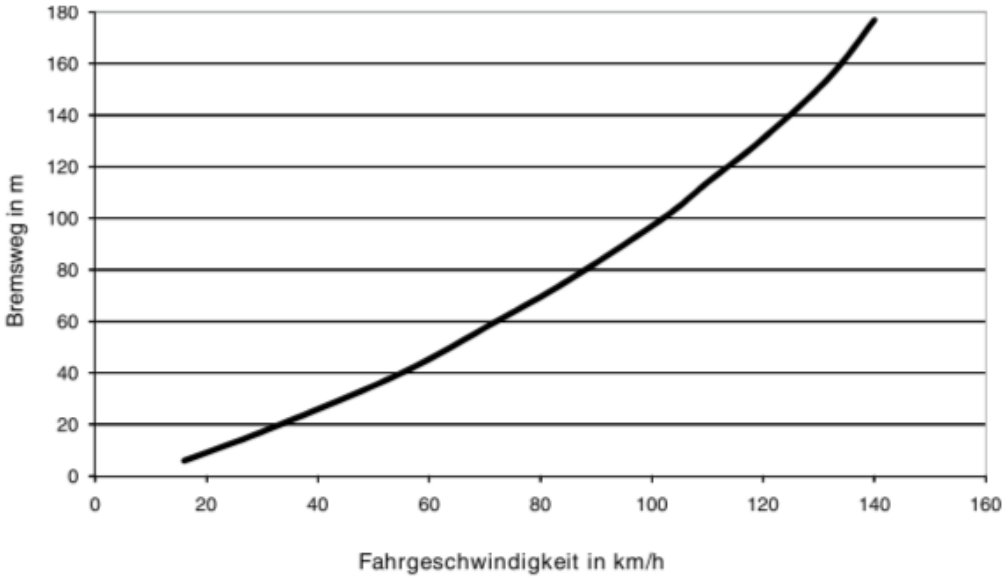


Bild 2.6: Zusammenhang zwischen Bremsweg und Fahrgeschwindigkeit [16]

Assistierenden Lichtverteilung

Die Verbesserung auf Basis der assistierende Lichtverteilung, sind im Grunde genommen Netzwerke aus Sensoren und diese Sensoren haben als Aufgabe, zeitliche und räumliche Erfassung des Verkehrsraumes, um Objekte schneller zu erkennen und zu klassifizieren und die Informationen über die Winkelpositionen der Objekte in horizontaler und vertikaler Richtung und Abstände der Objekte zum eigenen Fahrzeug herauszufinden. Ziel ist es, die Aufmerksamkeit der Fahrer zu wecken, das sogenannte Prinzip des Markierungslichtes. Die Abbildung 2.7 stellt das Markierungslicht dar. [17]

Bild 2.7: Markierungslicht [16]

2.2.2 Frontalkollisionsschutzsysteme

Längsverkehrsunfall ist die zweit häufigste Unfallart mit Schwerverletzungen und Tötungen und resultiert meistens aus einer kritischen Situation, mit unvorhersehbaren, aber vorhandenen Gefahren, die so genannten latenten Gefahren. Der Ablauf solcher kritischen Situationen, sind in der Abbildung 2.8 dargestellt. [17]Daraus folgen drei Strategien, wie man solche Situationen bewältigen kann, so dass der Unfall vermieden wird oder auf ein akzeptables Mindestrisiko reduziert wird. Dafür wurden entwickelt:

- Präventive Assistenz
- Reaktionsunterstützung
- Notmanöver

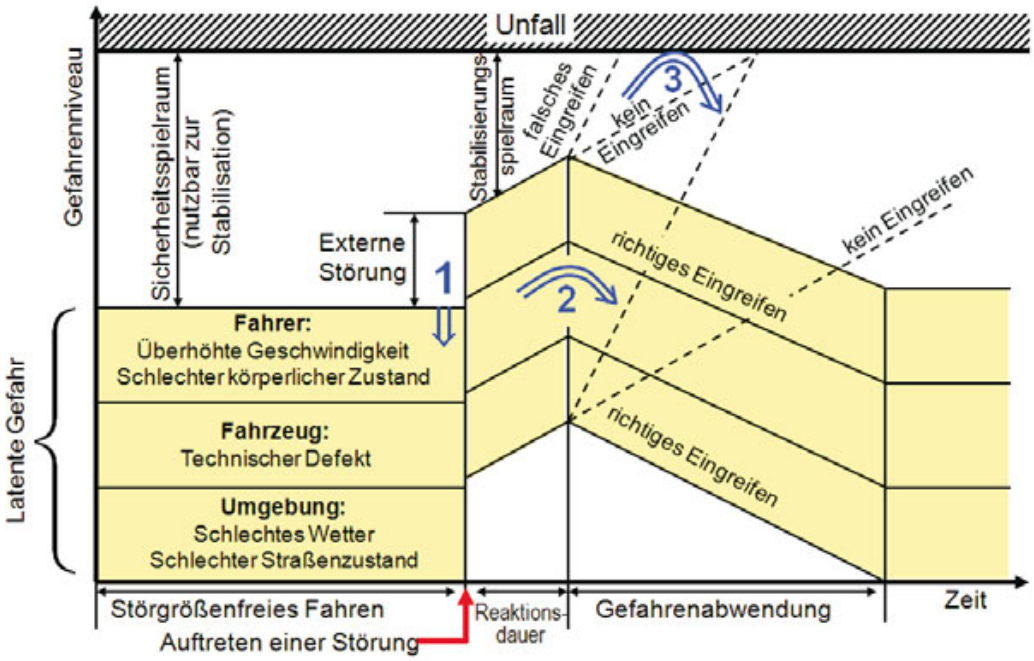


Bild 2.8: Assistenzstrategien in kritischen Verkehrssituationen [17]

Präventive Assistenz Zwei Möglichkeiten lassen sich hier Ableiten. Einerseits wird versucht, die latenten Gefahren herabzusetzen um die Eintrittswahrscheinlichkeiten in eine kritische Situation zu geraten stark zu verringern. Andererseits beim Eintritt der kritischen Situationen einen größere Handlungsspielraum für den Fahrer zu schaffen oder die Fahrerfähigkeiten zu erhöhen, um die Situation erfolgreich zu bewältigen. [17]

Reaktionsunterstützung

Unter Reaktionsunterstützung versteht man, die Aufmerksamkeitserregung, Situationsklärung und Unterstützung der Eingriffe. Die Aufmerksamkeit auf eine Situation zu erzielen, erfolgt durch Wahrnehmung, mithilfe von Warnsystemen. Hier es ist zu betonen, dass eine auditive Warnung effektiver ist und dementsprechend eine sehr hohe Aufmerksamkeit als eine visuelle Warnung bietet. In solchen kritischen Situationen gibt es zwei hauptsächliche Reaktionen zur Vermeidung des Unfalls, und zwar Ausweichen oder vor dem Hindernis Anhalten. [17]

Notmanöver

Unter Notmanöver versteht man ein autonomes Notmanöver und ist bezeichnet als ein „harter Eingriff in der letzten Sekunde vor einem vorhergesagten Aufprall, um den Schaden abzuwenden oder zu verringern.“ [17]

2.2.3 Lane Departure Warning

Unfälle mit schwerer Verletzung oder Tötungen kommen häufiger durch Spurwechsel oder durch ein unbewusstes Verlassen des Fahrstreifens und repräsentieren mehr als ein Drittel aller tödlichen Unfälle. In diesem Zusammenhang wurde ein Spurhalt Assistentensystem entwickelt; meist in Form von Warnung bei unbewusstem Verlassen von Fahrstreifen. [17]

Fahrstreifenerkennungssysteme Die Umweltbedingungen auf der Fahrbahn sind ein großer Störfaktor, wenn es um Effektivität eines Fahrstreifenerkennungsystems geht. Dennoch ist es Vorteilhaft, solche Systeme zum Einsatz zu bringen, weil diese im Vergleich zum Menschen, nie abgelenkt oder müde werden. Sie bestehen aus [17] :

- Sensor zur Erfassung der Umgebung
- Algorithmen, die die Fahrbahnmarkierungen erkennen
- Entscheidungseinheit zur Warnung
- Aktor zur Erkennung des Warnsignals oder zur Durchführung eines Lenkeingriffs.

Fehler solcher Systeme entstehen meistens aus einer nicht zusehenden Fahrmarkierung, die entweder durch Gras, Sand oder Schnee verdeckt werden.

2.2.4 Lane Departure Warning-System(LDW)

Die LDW dient dazu, den Fahrer vor einem unbewussten Verlassen des Fahrstreifens zu warnen. Zwei Kriterien helfen dabei, dies zu gewährleisten, und zwar die DLC-Kriterien (Distance to Line Crossing) und die TLC-kriterien (Time to Line Crossing). [18] Das DLC-Kriterium basiert nur auf dem lateralen Versatz des Fahrzeugs und zieht nur den Abstand des Rades zum Fahrstreifen in Betracht, ohne in die Zukunft zu prädizieren. [18]
Das TLC Kriterium legt eine vordefinierte zeitlichen Grenze fest, berechnet dann die Zeit bis zum Überfahren der Markierung und warnt den Fahrer beim Unterschreiten der Grenzwerte. Die Warnungen können sowohl visuell, akustisch oder haptisch erfolgen, um dem Fahrer zu ermöglichen, schnell auf eine Warnung zu reagieren. [18]
Eine visuelle Warnung kann durch ein leuchtendes Symbol oder eine Einblendung realisiert werden. Bei einer akustischen Warnung kann ein Signalton über einen Lautsprecher ausgegeben werden. Und Haptische Warnungen “ können über ein Lenkradmoment (z. B. Lexus), ein vibrierendes Lenkrad (z. B. Audi, BMW) oder einen bzw. zwei Vibratorelemente im Fahrersitz (z. B. Citroën) erfolgen.“ [18]

2.2.5 Advanced Lane Departure WarningSystem (ALDW)

Generell dient ein LDW-Systems dazu, den Fahrer beim ungewollten Verlassen des eigenen Fahrstreifens zu warnen. Die ALDW ist eine verbesserte Version von LDW und erfüllt einige zusätzliche Anforderungen, wie zum Beispiel das transparente Arbeiten des Systems, sodass die Warnung dann rausgeben wird, wenn eine Markierung am Boden überfahren wird. Diese soll dem Fahrer nichts Vorschreiben und auch nicht irritieren. [18] eine ALDW sieht so aus : „überfährt das Fahrzeug die Fahrstreifenmarkierung ohne gesetzten Fahrtrichtungsanzeiger, wird eine Warnung ausgegeben. Auf einer kurvigen Landstraße wird jedoch ein Fahrer hin und wieder die Fahrstreifenmarkierung überfahren, obwohl er weder abgelenkt noch unaufmerksam ist.“ [18] Bei Überholungen oder Spurwechseln wird der Fahrtrichtungsanzeiger ganz selten gesetzt.

2.2.6 Lane Keeping Support

Sie ist eine funktionale Erweiterung eines Lane Departure Warning- Systems. Ein LKS-System misst die Position der Fahrzeuge, in Relation zur Fahrstreifenmitte. Damit unterstützt sie den Fahrer die Spur zu halten. Beim Verlassen des Fahrstreifens korrigiert das System kurz vor dem Verlassen des Fahrstreifens die Position, indem es gezielten Lenkeingriff übernimmt. Jedoch ist die Kontrolle jederzeit dem Fahrer überlassen. Dieser systematische und gezielte Lenkeingriff dient meisten in einer Notsituation den Fahrstreifen verlassen zu können und eine Notfallsituation zu vermeiden. [18]

2.2.7 Lane Departure Prevention (LDP)

LDP ist eine Erweiterung für Advanced Lane Departure WarningSystem. Hier wurden viel mehr Werte auf dem Sicherheitssystem festgelegt. [18] Kommt das Fahrzeug vom Fahrstreifen ab, wird der Fahrer erst gewarnt und wenn keine Aktivität des Fahrers vorgesehen ist, wird das Fahrzeug nach dem Verlassen des Fahrstreifens aktiv zurückgelenkt. Dies geschieht durch den Bremseingriff mittels eines ESP-Systems, dass es ermöglicht, jedes Rad einzeln abzubremsen. [18] und das Fahrzeug durch einen einseitigen Bremseingriff zurück in den Fahrstreifen zu führen. [18]

2.3 Vorhersagende Spurwechselnde Modelle “Lane changing Model“

Spurwechsel ist definiert als ein Fahrmanöver, dass ein Fahrzeug von einer Spur zur anderen bewegt und die beiden Spuren in derselbe Fahrtrichtung gerichtet sind. Dieses Fahrverhalten

hängt auch von der Kinetik mehrerer Fahrzeuge ab. Man sollte in diesem Prozess die Fahrzeuge, sowohl auf der ursprünglichen Spur, als auch auf der angrenzenden Spur, beachten [2]. Die Anzahl der Spurwechselunfälle in den letzten Jahren beträgt jährlich zwischen 240.000 und 610.000 in den USA, was zu mindestens 60.000 Verletzungen führt.
[19]Zahlreiche Methoden und Modelle wurden in verschiedenen Bereichen vom Fahrzeug vorgeschlagen. (siehe Tabelle 2.3).

Olson führt umfassende Forschungen zum Spurwechsel Verhalten und auch zahlreiche Analysen über seine Spurwechsel Szenarien. Gipps [20] und Ahmed nutzen das „qq gap acceptance “Modell zur Darstellung des Längsverhaltens vor dem Spurwechsel. Kesting et al. schlägt vor, dass während des Spurwechsels das Bremsen vermieden werden soll. Oliver und Pentland verwendeten das Hidden-Markov-Modelle (HMM) um den Spurwechselzustand zu erkennen. Sie verwendeten verschiedene Funktionen, um Spurhaltungs- und Spurwechselverhalten zu klassifizieren. In [21] verwendeten McCall et al. die Breite Position und die Kopfbewegung, um sich auf die Spurwechselabsicht festzulegen. Ihre Studie dient dazu, eine freie Bahn zum Spurwechseln zu erkennen und mögliche Kollisionen vor den Spurwechsel Manövern voherzusagen. In [22] Pentland und Liu entwickelten ein Hidden-Markov-Modell mithilfe eines dynamischen Modells, um das Fahrverhalten zu erkennen und nachfolgende Aktionen der Fahrer vorherzusagen. In [23] verwendete Salvucci kognitive Architekturen, um die Beziehung zwischen dem visuellen Teil und des toten Winkels, während des Spurwechsels zu analysieren. In [24],verwendet Kim et al. das stetige Polynommodell zur Darstellung verschiedener Stufen des Spurwechselverhaltens mithilfe unterschiedlicher Regressionsmodelle und analysiert das Verhalten verschiedener Fahrer. . . etc.
In diesem Abschnitt wird es detailliert erläutert, wie das „ Lane changing model(LCM)“ aufgebaut ist. Es wird hier in zwei Hauptverhalten unterschieden. Die „ mandatory lane changing (MLC)“ und „ discretionary lane changing (DLC)“.Wenn ein Spurwechsel erforderlich ist, aufgrund eines wegfallen von Fahrstreifen oder das Platz machen für den Verkehr auf einer Auffahrt Spur, wird dieses als obligatorischer Spurwechsel (MLC) bezeichnet. Andererseits ein Spurwechsel, um die Fahrbedingungen zu verbessern, also aufgrund eines langsamen Fahrers oder aufgrund der schlechten Fahrbahn, wird als „ discretionary lane changing (DLC)“ bezeichnet. [19] Der Vorgang von Lanchaging besteht aus drei Stufe:

- Entscheidungsstufe (anhand von DLC or MLC)
- Wahl der Zielspur
- Annahme von freien Lücken

Tabelle 2.3: Spurwechselnde Modelle [19]

Autoren	Modelle oder Methoden	Bereiche
Olsen	Statistics analysis of natural Lane-change data	Analyze natural lane-change behavior
Gipps	Hierarchical decision method and-longitudinal dynamic model	Model lane-change decision behavior
Ahmed		
Kesting		
Penttland	HMM and dynamic models	Model and predict driver behavior (including Lane-change behavior)
Salvucci	Cognitive architectures (ACR-T)	
Kim	PWP	
Sekizawa	SS-ARX	
Tan	A drivers steering model based on the target angle error	
Oliver	HMM	Recognize lane-change state
Mandalia	SVM	
Salvucci	A computational model	
McCall	Sparse Bayesian learning	Infer lane-change intent
Enke	A sinusoidal model,	Generate lane-change trajektory
Choven	A sinusoidal model and a trapezoidal model	
Nelson , Papadimitri	A fifth-order polynomial model	
Choven	A sinusoidal model and a trapezoidal model	Analyze safe and unsafe lane-changes
Bascunana	a longitidunal dynamic models	
Jula	a longitidunal dynamic models with a latitudinal Sinusoidal model	
Smith	satatistics analysiy of the natural lane chnge data	
Schubert, Broadhurts	Probability model of lane-change maneuver	

2.3.1 Struktureller Aufbau von dem Lane changing Model

Zwei Unterschiedliche Verhalten werden hier abgebildet. Ein visueller Teil und ein latenter Teil. [25] Die in Abbildung (2.9) abgebildete ovale stellt den latenten Teil dar. Unter latent bezeichnet man einen Entscheidungsprozess. Das Rechteckige stellt wiederum den visuellen Vorgang dar, zum Beispiel das Wechseln auf eine linke oder rechte Spur. Wenn man mit MLC anfängt, bedeutet das, dass man sich im Bereich des obligatorischen Spurwechsel Vorganges befindet. Das Verhältnis zwischen dem zurückgelegten Weg und dem Punkt, wo ein Spurwechsel erfolgen sollte. Die Art des Fahrzeugs (LKW, Bus oder PKW) und der Anzahl von Fahrstreifen, die überquert werden müssen, um auf die gewünschten Spuren zu kommen, sind die Einflussfaktoren des MLC. Zum Beispiel der Fahrer fängt die Lane changing Manöver frühzeitig an, wenn er viele Fahrstreifen überqueren muss. Der Linke Zweig entspricht dem DLC Vorgang (Abb. 2.9). Viele Faktoren beeinflussen diese Entscheidung, und zwar der Geschwindigkeitsunterschied zwischen dem Nachbarfahrzeug, gewünschte Fahrgeschwindigkeit und die Dichte der verschiedenen Fahrbahnen. Die unterste Ovale stellt die Phase der Annahme von freien Lücken dar. Es ist festzustellen, dass die erforderlichen Lücken abhängig von der Geschwindigkeit sind.

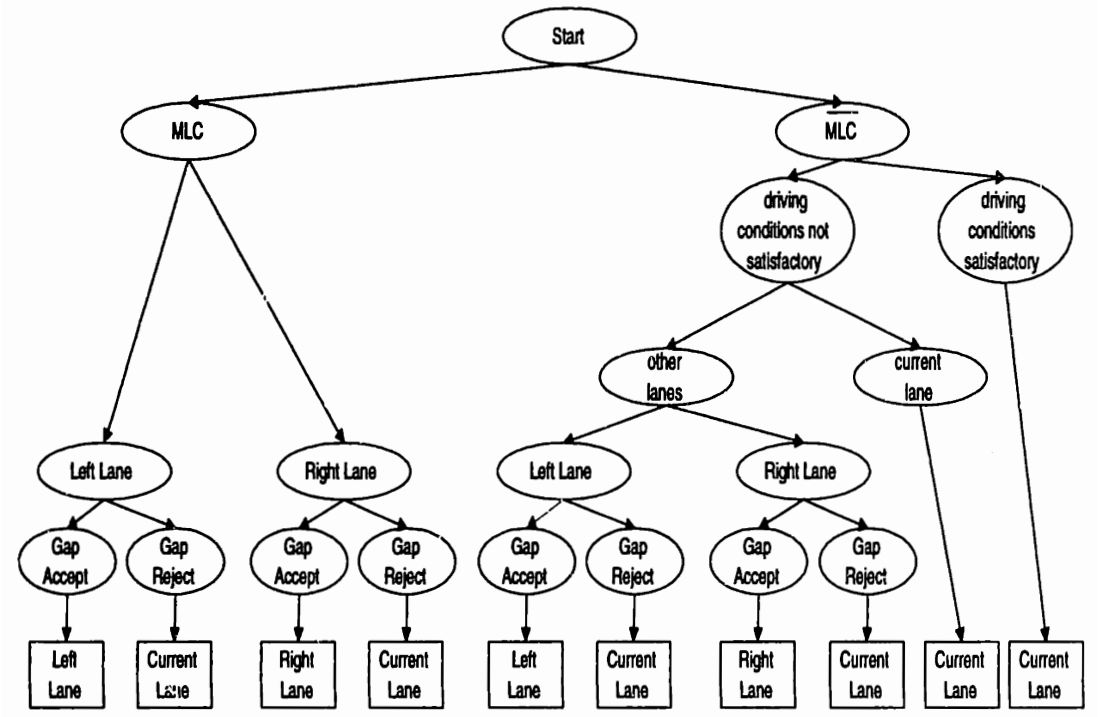


Bild 2.9: Struktureller Aufbau von Spurwechsel [25]

2.3.2 Spur Auswahl Modelle

Für die Entscheidung, ob es um MLC oder DLC Spurwechsel geht, wird mit einem diskreten Modell behandeln, zum Beispiel das binäre Logik Modelle basierend auf dem Zufall Ausdruck. [25]die Wahrscheinlichkeit, dass ein Fahrer n auf einer Zeit Punkt t für eine mandatory lane changing auswählen unter Rücksicht von einer gewisse Menschen spezifische Zufallszahl ν_n, ist gegeben durch :

$$P_t(MLC|\nu_n)=\frac{1}{1+e^{(-X_n^{MLC}(t)\beta^{MLC}-\alpha^{MLC}\nu_n)}} \tag{2.1}$$

wobei:
X_n^{MLC} :
Erklärende vektoriellen Einfluss Variablen , die die MLC Entscheidung beeinflussen β^{MLC}:
Vektorieller Parametern α^{MLC}: ν_n Parametern
ν_n : Menschen spezifische Zufallszahl die Normal verteilt sind

ν_n erfasst die Korrelation zwischen verschiedene Beobachtung für eine gegebene Fahrer. Je größer die Produkte zwischen ν_n Parametern, desto größer ist die Wahrscheinlichkeit, dass der Fahrer frühzeitig für die MLC entscheidet. Die binäre Entscheidung ob der Fahrer auf die Bedingungen der Fahrbahn zufrieden ist oder nicht, wird durch Folgende Binäre logische Modelle gegeben:

$$P_t(DCNS|\nu_n)=\frac{1}{1+e^{(-X_n^{DCNS}(t)\beta^{DCNS}-\alpha^{DCNS}\nu_n)}}$$

(2.2)

DCNS steht für Driving Condition not Satisfactory. Im Fall von DCNS, also wenn die Fahrtbedingungen nicht gut ist und der Fahrer seine aktuelle Spur mit den restlichen Spuren vergleicht. Das Lane changing Modell dient dazu, eine bessere Spur zu identifizieren und in die nächste Stufe zu kommen, wo der Fahrer nur dann eine freie Lücke sucht. [25]

2.3.3 gap acceptance model von Ahmed

Das Modell erfasst und bewertet die Lücken (lag und Lead gap) als akzeptabel oder nicht akzeptabel. Es wird angenommen, dass die Fahrer nur die benachbarten Lücken berücksichtigen. Eine benachbarte Lücke ist definiert als die Lücke zwischen dem Führungsfahrzeug (lead vehicle) und dem Nachlauffahrzeug (lag Vehikel) in der Zielspur (siehe Bild 2.10). Für das Zusammenführen in eine benachbarte Spur ist eine Lücke nur dann akzeptabel, wenn bLead gap und lag gap akzeptabel sind. [25]

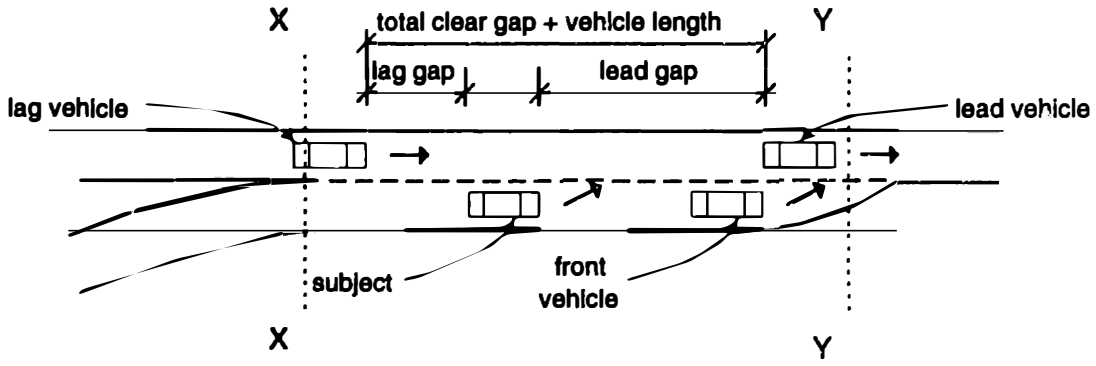


Bild 2.10: Darstellung von Fahrbahn [16]

Jeder beliebige Fahrer hat eine akzeptabel Lead and lag gap, die als Kritisch gap bezeichnet wird. Der Kritische Gap für ein gegeben Fahrer n auf eine Zeit t ist definiert als

$$G_n^{cr,g}(t)=e^(-X_n^{X_n^g}(t)\beta^g + \alpha^g v_n + \epsilon_n^g(t)) \tag{2.3}$$

wobei,
$g \in$ {lead, gap},
α^g : Parameter für v_n mit $g \in \{lead, gap\}$,
$\epsilon_n^g(t)$: generiert Zufallszahl mit Wertebereich zwischen g, t, n Die Exponentielle in das Formelle deutet hin, dass die kritische Werte im Positive sein wird.v_n beschreibt die Korrelation zwischen kritische Lag und die Kritische Lead für eine gegebene Fahrer . Angenommen $g, n(t)$ standard normal verteilt sind und die Kritische gap Abstand Logarithmisch normal verteilt ist, dann die bediente Wahrscheinlichkeit eine akzeptabel Lücke ist gegeben durch:

$$P_t(gapAcc|v_n) \quad = P_t(\text{lead gap acceptale}|v_n)P_t(\text{lag gap acceptable}|v_n) \tag{2.4}$$

$= \mathrm{P}(G_n^{lead}(t) > G_n^{cr,lead}(t)|\nu_n) P(G_n^{lag}(t) > G_n^{cr,lag}(t)|\nu_n)$
$= P(\ln(G_n^{lead}(t)) > \ln(G_n^{crlead}(t))|\nu_n) P(\ln(G_n^{lag}(t)) > \ln\ln(G_n^{lead}(t))|\nu_n)$
$= \phi(\frac{\ln(G_n^{lead}(t)) - X_n^{lead}(t)\beta^{lead} - \alpha^{lead}\nu_n}{\sigma_{\epsilon n}^{lead}}) * \phi(\frac{\ln(G_n^{lag}(t)) - X_n^{lag}(t)\beta^{lag} - \alpha^{lag}\nu_n}{\sigma_{\epsilon n}^{lag}})$
wobei,
lead gap: $G_n^{lead}(t)$
lag gap: $G_n^{lag}(t)$
ϕ: Die Kumulative Verteilungsfunktion ist

2.4 Relevante Methoden

2.4.1 Angewandte Lineare Regression

Ein Regressionsmodell ist ein quantitatives Analyse Tool, in dem das Verhalten einiger Variable durch andere Variablen erklärt wird. Eine Regressionsmodellgleichung hat die Form einer mathematischen Funktion, die die Beziehung zwischen einer abhängigen Variablen und einer oder mehrere erklärenden Variablen quantitative beschreibt. Eine theoretische Notation, die keine spezifische Funktion (mathematisch) voraussetzt, sind in der Form:

$$Y = f(X_1, X_2, ...X_i, \epsilon) \tag{2.6}$$

wobei,
Y : Abhängiger Variabel
$X_1, X_2, ...X_i$: Erklärende Variablen
ϵ: Residuen aus dem Modell

An dieser Stelle liegt der Fokus auf dem linearen Regressionsmodelle mit einer Gleichung, die wahrscheinlich die am häufigsten angewandte ist. Die primären Elemente solcher Modelle sind:

- **Koeffizienten** (Strukturparameter) des Modells: Achsenabschnitt, der im Fall von Modellen mit einer einzigen erklärenden Variablen als Schnittpunkt einer Regressionslinie mit einer vertikalen Achse des scatterplots interpretiert wird (d.h. der erwartete Wert der abhängigen Variablen bei Nullwert der erklärende Variable)

- **Steigungskoeffizienten**, die einzelne erklärenden Variablen zugeordnet sind und die Richtung und Größe der Auswirkung einer bestimmten erklärenden Variablen auf die abhängige Variable quantifizieren.

- **Standardfehler** einzelner Strukturparameter, die die erwarteten Bereiche bestimmen, in die die wahren Werte einzelner Koeffizienten (die möglicherweise nicht genau den geschätzten entsprechen) fallen.
- **T-Statistiken einzelner Koeffizienten**, berechnet durch Division des Werts eines gegebenen Koeffizienten durch seinen jeweiligen Standardfehler, der zur Bewertung der statistischen Signifikanz einzelner Koeffizienten (und damit auch der statistischen Signifikanz einzelner erklärender Variablen) herangezogen wird .
- **Bestimmungskoeffizient**, auch R-Quadrat R^2genannt, misst die Anpassungsgüte des geschätzten Modells an die tatsächlichen Daten. R^2 wird mit dem folgenden Formel bestimmen:

$$R^2 = 1 - \frac{\sum_{t=1}^{n} e_t^2}{\sum_{t=1}^{n} (y_t - \bar{y})^2} \tag{2.7}$$

wobei,

n: Anzahl der Beobachtungen, die zur Schätzung eines bestimmten Modells ,
y_t: tatsächliche Beobachtungen der abhängigen Variablen,
e_t: Residuen des Modells,
$\bar{y}$geschätzte Beobachtungen der Abhängigen Variable

Eines der Grundprinzipien der Modellbildung besteht darin, nur diejenigen erklärenden Variablen in das Modell aufzunehmen, die statistisch signifikante Beziehungen zur abhängigen Variablen aufweisen. Die Schätzung eines Regressionsmodells erfordert daher die Bewertung der statistischen Signifikanz seiner einzelnen erklärenden Variablen. Das am häufigsten angewendete Werkzeug zur Überprüfung der statistischen Signifikanz einzelner erklärender Variablen ist ein Test, der am besten für diese Variablen berechnet wird.

2.4.2 Regression Logistik

Die logistische Regression ist eine der am häufigsten verwendeten Klassifizierungsmethoden und funktioniert gut, wenn Daten linear trennbar sind.Es gibt viele wichtige Anwendungsfälle, bei denen die abhängige Variable nicht kontinuierlich ist, z. B. Spurwechsel oder keine Spurwechsel, Kundenabwanderung oder keine Kun- denabwanderung. Angenommen, wir untersuchen, ob ein Kunde abwandert (y =0),oder läuftzeichnet wir die Wahrscheinlichkeit, dass eine Kunde abwandern als P = P (Abwanderung) und die mögliche erklärende Variablen sind:

Rechnungen, Rechnungsbetrag und Anzahl von Tagen, in den keine Rechnungsausgleiche stattgefunden haben.

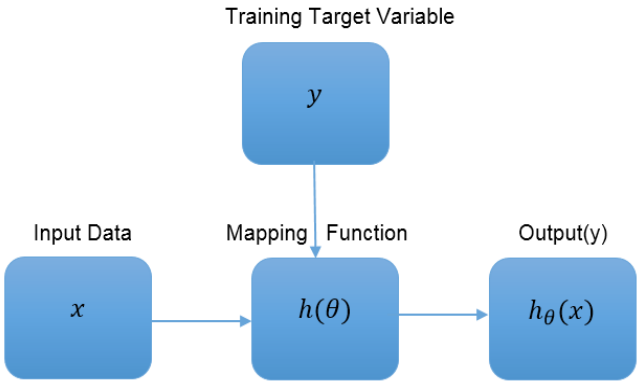


Bild 2.11: Schematische Modellierung von Regression Logstik [16]

Regression Logistik ist anzuwenden, wenn folgende Voraussetzungen erfüllt sind:

- Die Beziehung zwischen der unabhängigen Variablen und der abhängige variablen ist nicht linear.
- Der Irrtum ist nicht normal verteilt.

$$\log_i t(p) = \log(\frac{p}{1-p}) = \theta_0 + \theta_1 * x_1 + \theta_2 * x_2 + \theta_3 * x_3 ... \tag{2.8}$$

2.4.3 Prinzipal Komponente Analysis

Die Hauptkomponentenanalyse (PCA) bildet die Grundlage für die multivariaten Daten Analyse. [26] PCA liefert einen Datensatz in Form auf einer Datenmatrix Xals Produkt zweier kleiner Matrizen T und P'. Diese Matrizen T und P 'erfassen die wesentlichen Datenmuster von X, wobei die Darstellung der Spalten von T beinhalte die Mehrheit der Informationen die in X steckt und P´ beinhalte die restlichen Information. Die beobachteten Merkmale werden durch eine lineare Transformation in einer unkorrelierte überführt. Ziel ist der PCA, die komplizierten Beziehungen zwischen den Daten auf einfache Form zu reduzieren. [26]

2.4.4 Entscheidungsbäume

Entscheidungsbäume sind eine Methode der Aufteilung von Objekten anhand geeigneter Merkmale in Gruppen im Hinblick auf eine vordefinierte Zielgröße. [26] Es wird in zwei Varianten unterteilt: Klassifikationsbäume und Regressionsbäume. Klassifikationsbäume werden bei nominal skalierten Variablen als abhängige Zielgröße verwenden, während bei Regressionsbäumen eine quantitative Variable als abhängige Zielgröße genutzt wird. [26]

3 Statistische Analyse der HighD-Messdaten

3.1 Beschreibung der Messdaten

Für den HighD-Datensatz wurde der Verkehr auf deutschen Autobahnen mit Drohnen erfasst. Um die Trajektorien aus den Aufzeichnungen zu extrahieren, wurden modernste Computer Vision-Algorithmen verwendet. Unter Verwendung neuronaler Netze wurden die Fahrzeuge erkannt und lokalisiert. [27]Um aus diesen Erkennungen glatte Flugbahn zu erhalten, wurden die Fahrzeuge über die Zeit verfolgt und ihre Bewegung wird mithilfe der Bayes'schen Glättung korrigiert. Der HighD-Datensatz wurde an insgesamt sechs verschiedenen Standorten in der Nähe von Köln aufgenommen. Die Standorte variieren je nach Anzahl der Fahrspuren und Geschwindigkeitsbegrenzungen. [27]Die Aufnahmen beinhalten sowohl leichte als auch starke Verkehrsaufkommen.

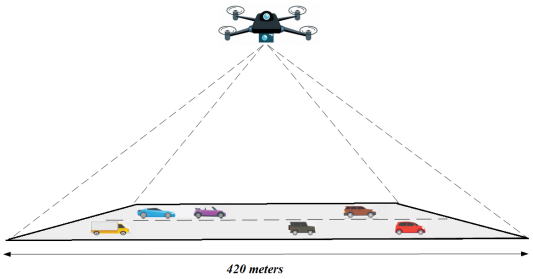


Bild 3.1: Drohnen Messung [28]

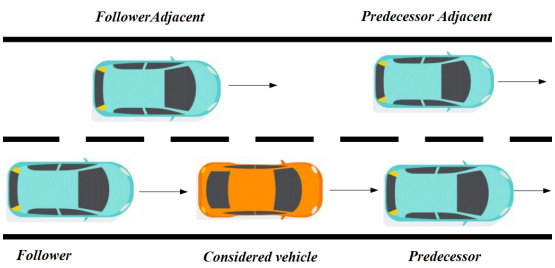


Bild 3.2: Drohnen Messung

Bestehende Datensätze sind normalerweise ungenau, da einige Fahrzeuge nicht richtig oder nicht genau erkannt werden. Wenn mehrere Kameras gleichzeitig verwendet werden, können auch bei den Übergängen Probleme auftreten. [29] [30]Im Gegensatz dazu wird für den HighD-Datensatz nur eine einzige 4K-Kamera mit sehr hoher Auflösung verwendet. In Kombination mit modernsten Algorithmen für maschinelles Lernen ergibt sich ein typischer Positionsfehler der erkannten Fahrzeuge von 10 cm. [27]
Für jedes Fahrzeug wird eine eindeutige ID für die Verfolgung zugewiesen, und seine Geschwindigkeit wird über die Zeit geschätzt (besser: approximiert). Darüber hinaus stellt diese Arbeit einen umfangreichen naturalistischen Fahrzeugtrajektionsdatensatz von deutschen Autobahnen mit dem Namen HighD zur Verfügung.

3.1.1 explorative Datenanalyse

Graphische Darstellung der Variablen

Die Abbildungen (3.2 und 3.3) sind die paarweise Darstellung von den 25 Variablen bezüglich der Manöver. In Bild 3.2 sind in orangener Farbe und pro Kästchen das Verhältnis von zwei Variablen bezüglich des Spurwechsels auf der linken Seite und in Blau das Spurhaltende auf der rechten Seite. Es ist anhand der Form der Punktwolken festzustellen, dass einige Variablen stark miteinander korrelieren, manche ganz wenig und manche überhaupt nicht korrelieren. Beispielsweise wird der Graph (TimeGapred ,dispred) betrachtet, merkt man sofort dass, die beide Variabel, bezüglich der beiden Manövern im Vergleich zum Rest, eine starke Korrelation aufweisen. Ferner weisen die Paare (Xvelocity,diffspeedpred) eine schwache Korrelation auf, weil die Punktwolke eine gewisse Linearität verliert. Darüber hinaus sind die Paare (yvelocity,x Accelarati- on) interessant, da die Yvelocity keine Auswirkung auf das Spurverhalten hat. Ebenfalls hat Xacceleration keinen Einfluss auf einen Spurwechsel.

Bild 3.3: Paarweise Darstellung der Variable LCL-LKR

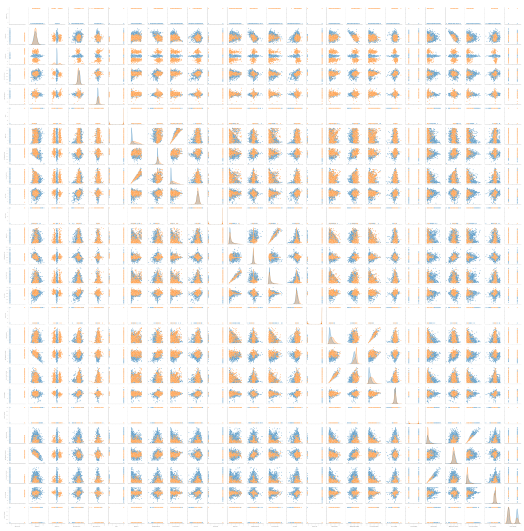


Bild 3.4: Paarweise Darstellung der Variabel LKL-LCR

Lage Maße der Variablen LKL-LCR

Die Lage Parameter, und zwar der arithmetische Mittelwert, die Standardabweichung, die Erste, zweite und das dritte Quantil für alle Variablen sind in Abbildung (3.4) dargestellt und umfassend im Anhang. Wenn man die Xvelocity in X-Richtung betrachtet, haben 25% der Fahrzeuge die Geschwindigkeit geringer als 30,18 Km/h, 50% kleiner als 33,09 Km/h (Me- diane) und 75% kleiner als 36,14 Km/h. Die maximale Geschwindigkeit liegt bei 59,1 Km/h, die minimale Geschwindigkeit liegt bei 17,71 Km/h mit einer mittleren Geschwindigkeit von 33,28Km/h. Im Anhang dieser Arbeit befindet sich die vollständigen Lage Maße für alle Variablen.

	maneuver	xVelocity	yVelocity	xAcceleration	yAcceleration	\
count	6822.000000	6822.000000	6822.000000	6822.000000	6822.000000	
mean	2.090003	33.288665	-0.014592	0.019008	-0.005141	
std	0.286207	4.657603	0.297218	0.432222	0.061745	
min	2.000000	17.710000	-1.700000	-3.500000	-0.750000	
25%	2.000000	30.180000	-0.070000	-0.220000	-0.040000	
50%	2.000000	33.090000	-0.010000	0.040000	-0.010000	
75%	2.000000	36.140000	0.050000	0.300000	0.030000	
max	3.000000	59.100000	1.730000	1.840000	0.510000	

	class	distPred	diffSpeedPred	timeGapPred	accPred	...
count	6822.000000	6822.000000	6822.000000	6822.000000	6822.000000	...
mean	0.982556	80.804795	-0.050431	2.360928	0.022962	...
std	0.130927	83.457110	3.311015	2.399467	0.414062	...
min	0.000000	2.320000	-28.030000	0.090000	-2.910000	...
25%	1.000000	27.200000	-1.100000	0.860000	-0.210000	...
50%	1.000000	45.275000	0.000000	1.370000	0.030000	...
75%	1.000000	99.195000	1.150000	2.820000	0.280000	...
max	1.000000	406.710000	16.450000	16.000000	1.580000	...

Bild 3.5: Lage Parameter

Lage Maße der Variablen LKR-LCL

Die Abbildung (3.5) stellt die Lage Maße für der Fall LKR-LCL da. Die Betrachtung der Xvelocity ergibt folgendes:25% haben die Geschwindigkeit in X-Richtung kleiner als 23,64 Km/h; 50% haben die Geschwindigkeit in X-Richtung kleiner als 24,96Km/h(Mediane)und 75% haben die Geschwindigkeit in X-Richtung kleiner als 29,75Km/h. Andererseits liegt die Maximale Geschwindigkeit bei 49,2Km/h; die minimale Geschwindigkeit liegt bei 18,01Km/h mit einer mittleren Geschwindigkeit von 26,80Km/h. Im Anhang befindet sich die Lage Maße für Alle Variablen vollständig. Anhand beider Beschreibungen ist es nachvollziehbar, dass die Fahrzeuge auf der linke Spur, schneller sind als auf der rechten Spur. In Folge wird quantitative untersucht, wie die Variablen miteinander zusammenhängen.

	maneuver	xVelocity	yVelocity	xAcceleration	yAcceleration	class	distPred	diffSpeedPred	timeGapPred	accPred	...
count	4874.000000	4874.000000	4874.000000	4874.000000	4874.000000	4874.000000	4874.000000	4874.000000	4874.000000	4874.000000	...
mean	1.242511	26.804040	-0.045375	0.005127	-0.024867	0.543291	89.616504	-0.021662	3.339690	-0.010599	...
std	0.817838	4.342169	0.160402	0.225267	0.097482	0.498173	67.214162	4.643871	2.488459	0.226725	...
min	1.000000	18.010000	-1.180000	-1.810000	-0.570000	0.000000	2.660000	-22.960000	0.120000	-2.290000	...
25%	1.000000	23.640000	-0.080000	-0.140000	-0.050000	0.000000	39.247500	-1.570000	1.520000	-0.140000	...
50%	1.000000	24.960000	-0.020000	0.010000	-0.010000	1.000000	68.565000	-0.070000	2.560000	0.000000	...
75%	1.000000	29.747500	0.030000	0.140000	0.020000	1.000000	121.700000	1.150000	4.457500	0.140000	...
max	4.000000	49.200000	0.870000	1.280000	0.470000	1.000000	398.250000	24.770000	16.140000	1.410000	...

Bild 3.6: Lage Parameter

Ziel

Anhand realen Daten und mit Hilfe von verschiedene Algorithmus werden die Manöver vorhergesagt und mit Hilfe von Netlog eine Vorhersagen simuliert.

3.1.2 Korrelationsanalyse

Die graphische Darstellung ist eine Veranschaulichung, wie stark die Variablen paarweise miteinander korreliert. Das schwarze Kästchen weist auf eine stark negative Korrelation hin. Das helle Orange eher auf eine starke positive Korrelation. Die genauen Werte der Korrelation sind im Anhang zu finden.

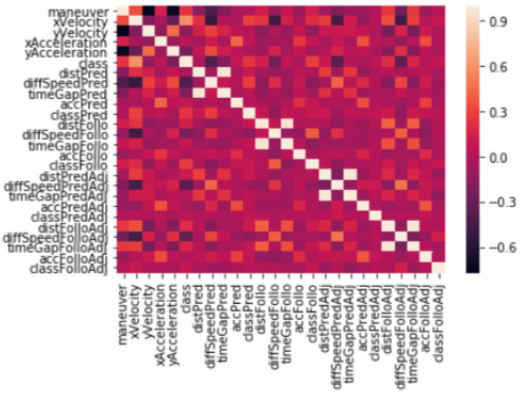


Bild 3.7: Korrelation Tabelle LCL-LKR

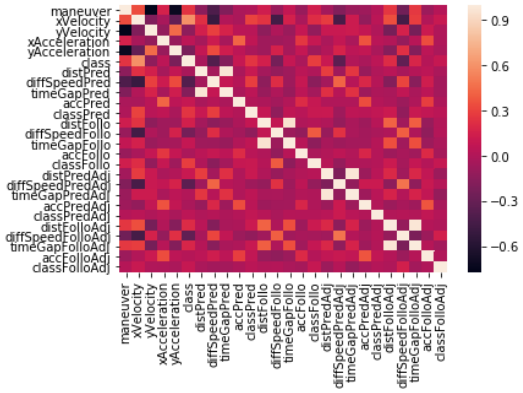


Bild 3.8: Korrelation Tabelle LCR-LKR

Die multivariabel Korrelationsanalyse zeigt, dass die Variablen paarweise korrelieren, aber es sei immer noch offen, wie die verschiedene Variablen die Manövern beeinflussen. Genauer gesagt, welche Variablen eine entscheidende Rolle spielen oder keine Rolle spielen, bei den verschiedene Manövern. Anhand dieser Information wird eine Hauptkomponentenanalyse durchgeführt, mit dem Ziel, die Datensätze besser zu verstehen.

3.1.3 Hauptkomponente Analyse der Messedaten

Die Hauptkomponentenanalyse (PCA) ist einer der am weitesten verbreiteten Methoden zur Datenexploration und –Visualisierung. [31]. PCA ist eine Projektion der Daten auf einen niedrige dimensionalen Raum und ist besonders geeignet, um Muster wie Cluster, Linien und Ausreißer in einem Datensatz zu visualisieren. Die Komponenten lassen sich auch graphische bestimmen. Als erste Komponente bezeichnet man die Hauptrichtung der Linie durch die Punktwolke mit dem Minimum der Summe der quadratischen Abstände von den Punkten zur Komponente. Mit Hilfe der Komponentenanalyse wird versucht mehr über die Datensätze zu erfahren. Im Grunde

genommen geht es um die Minimierung der Varianz zwischen verschiedene Variablen. Vorgehensweise:

- Standardisation von der Datensatz
- Berechnung der Kovarianz
- Berechnung des Eigenvektors
- Berechnung der PCA
- Reduktion der erklärende Variablen.

Die erste Hauptkomponente hat die größtmögliche Varianz und trägt den größten Anteil der Informationen. Die zweite Komponente wird orthogonal zur ersten Komponente gelegt/gesetzt und trägt den zweit größten Anteil der Information in den Datensätzen. Die andere Komponenten werden ebenfalls berechnet. Die Werte dieser neuen Variablen für die Beobachtungen werden als Faktor-scores bezeichnet Dies können als geometrisch Projektionen der Beobachtungen interpretiert werden.

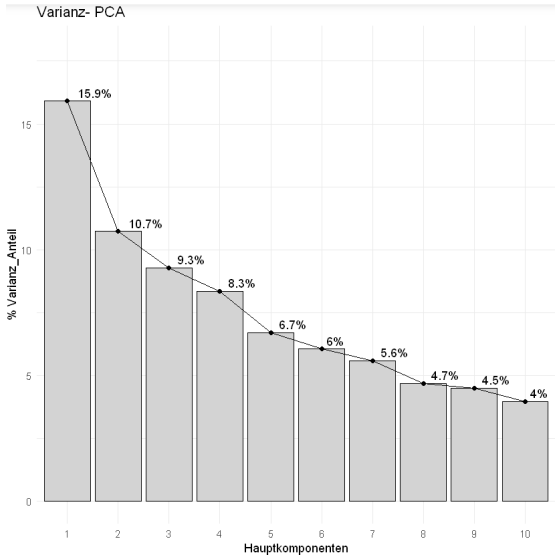


Bild 3.9: Prinzipal Komponente LCL-LKR

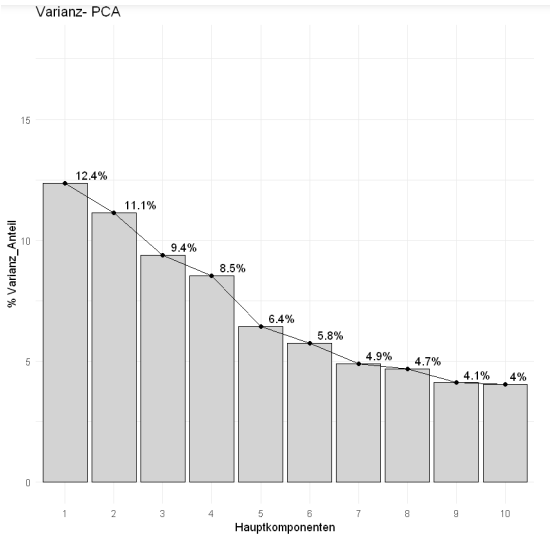


Bild 3.10: Prinzipal Komponente LKL-LCR

Das Ergebnis der kumulativen Varianzen (siehe Anhang) der PCA zeigt in beiden Fällen, dass sich 90% der Informationen in 16 Komponenten darstellen lassen. 38% der Informationen stecken in 3 Komponenten. Die PCA erlaubt uns zwar eine Reduktion der erklärenden Variablen von 26 Variablen auf 16 Komponenten. Wenn sich die Variablen nur auf die drei höchsten Varianzen beziehen ist eine Darstellung in einem 3-dimensionalen Raum möglich, jedoch gehen 60% der Informationen an dieser Stelle verloren.

Beitrag der jeweiligen Variablen zu einer Komponente

Der Wert eines Beitrags liegt zwischen 0 und 1 und für eine bestimmte Komponente ist die Summe der Beiträge aller Variablen gleich 1. Je größer der Wert des Beitrags, je mehr trägt die Variablen zur Komponente bei.

	Dim.1	Dim.2
maneuver	0.6462816537	-0.429540282
xVelocity	0.7904754772	0.184070103
yVelocity	-0.5402944395	0.338259176
xAcceleration	0.0447253785	-0.006910909
yAcceleration	-0.4811343489	0.379450645
class	0.6201844891	-0.015842403
distPred	0.0336229550	0.679819270
diffSpeedPred	-0.6259175248	0.178971718
timeGapPred	-0.1325977852	0.652153859
accPred	-0.0006761873	0.063487052
classPred	0.1767147036	0.239867869
distFollo	0.4045353761	0.048093259
diffSpeedFollo	-0.3633773502	-0.124471400
timeGapFollo	0.3319352738	0.050768000
accFollo	-0.0797953215	0.119805018
classFollo	0.2260799727	-0.011945219
distPredAdj	0.2177205360	0.717434788
diffSpeedPredAdj	-0.5732674762	-0.029421526
timeGapPredAdj	0.0947426184	0.705968064
accPredAdj	-0.1588305487	-0.010118429
classPredAdj	0.0214203569	0.047033091
distFolloAdj	0.5203645458	0.339305531
diffSpeedFolloAdj	-0.5660988423	0.016979427
timeGapFolloAdj	0.5243095975	0.325763523
accFolloAdj	-0.0504656392	0.213321380
classFolloAdj	-0.0171802564	-0.117882280

Bild 3.11: Beitrag der Variabel LCL-LKR

	Dim.1	Dim.2
maneuver	0.09646091	0.599747813
xVelocity	0.81039434	-0.007950761
yVelocity	0.02092117	-0.155948166
xAcceleration	-0.25222447	0.401610790
yAcceleration	0.05913447	0.141062821
class	0.09277139	-0.216848471
distPred	0.53025705	0.559428931
diffSpeedPred	-0.40611500	0.383626833
timeGapPred	0.42729462	0.598403189
accPred	-0.31811588	0.256977770
classPred	0.09433757	0.067744045
distFollo	0.62650525	-0.150296702
diffSpeedFollo	-0.25022864	-0.109428900
timeGapFollo	0.57655410	-0.134649932
accFollo	-0.02733246	0.411515996
classFollo	-0.02918851	-0.001866533
distPredAdj	0.25221219	0.536576166
diffSpeedPredAdj	-0.58161656	0.346074156
timeGapPredAdj	0.12647225	0.547041513
accPredAdj	-0.24759030	0.297551293
classPredAdj	0.09241447	0.239859403
distFolloAdj	0.16319618	-0.363489446
diffSpeedFolloAdj	-0.61990460	-0.126588662
timeGapFolloAdj	0.14731972	-0.337818560
accFolloAdj	-0.14653232	0.193951935
classFolloAdj	-0.04456413	-0.149333079

Bild 3.12: Beitrag der Variabel LKL-LCR

In dem Bild 3.10 ist festzustellen, dass die Geschwindigkeit der Autos auf der X-Richtung (xVelocity) bis zu 79% zu der erste Hauptkomponente beträgt, die Klasse der Fahrzeuge (LKW, PKW, Busse) trägt zu 62% für die erste Hauptkomponente und die diffSpeedPred (Geschwindigkeitsdifferenz zwischen Fahrzeugen) trägt auch zu 62% bei.Wenn die zweite Hauptkomponente betrachtet

wird, merkt man, dass distPredAdj zum 71% zu der zweite Hauptkomponente beiträgt, die timeGapPredAdj bis zu 70% und distPred bis zu 67% zu der zweite Komponente beiträgt. Für der Fall LKL-LCR trägt die Geschwindigkeit der Fahrzeuge auf der X-Richtung (xVelocity) zum 81% zu der erste Hauptkomponente bei, die disFollo der Fahrzeuge trägt zu 62% für die erste Hauptkomponente bei und die diffSpeedFolloAdj trägt zu 61,9% bei. Wenn man die zweite Hauptkomponente betrachten, trägt timeGapPred bis zu 59,8%zu der zweite Hauptkomponente bei und die distPred bis zu 55,9% der Zweite Komponente bei.

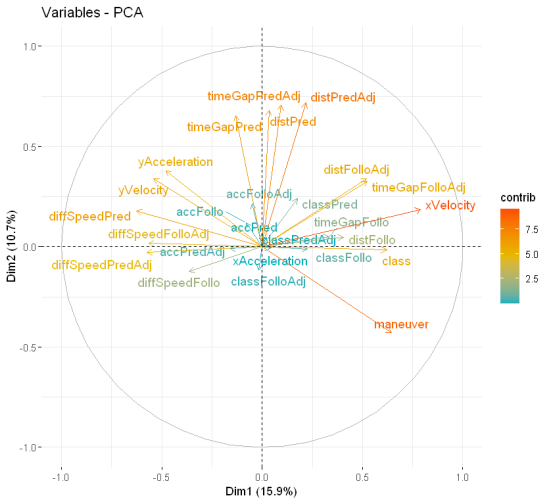


Bild 3.13: Korrelationskreis LCL-LKR

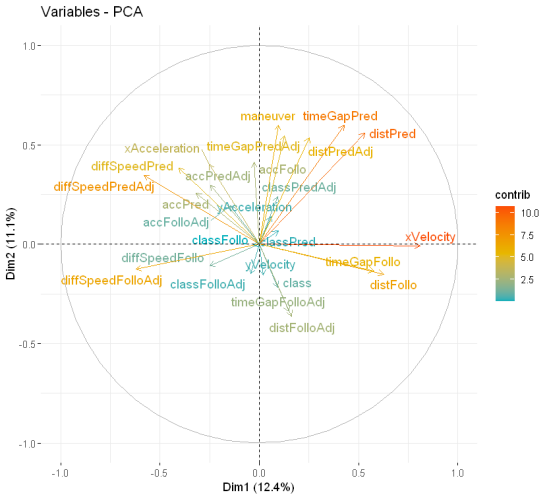


Bild 3.14: Korrelationskreis LKL-LCR

Korrelationskreis

Der Korrelationskreis hilft uns die Beiträge der Komponenten besser zu verstehen. Auf dem Bild 3.12 sind die Variablen die Stark zu einer Komponente beitragen zu sehen, die nur wenig zu der zweiten Komponente beitragen. Beispiele, die timeGapPredAdj, distpredAdj, timeGapPred tragen zur der zweite Komponente aber ganz wenig zu der erste Komponente bei, weil deren Projektion auf der erste Komponente gegen Null läuft. Die Variablen deren Projektionen gegen 1 oder -1 laufen, tragen zur Komponente bei. Die gleiche Analyse gilt auch für das Bild (3.13). Die PCA gibt folgende Schlussfolgerung, jedoch Hilfst es nicht weiter die Phänomen des Manövrierens besser zu beschreiben. Viele Variabel spielen eher eine Rolle, was das Spurwechseln Manövern angeht. Manche davon haben einen stärkeren Einfluss.

Beschreibung des Verhältnis der Variablen auf der Erste Hauptkomponente

Das Verhältnis lässt sich durch die Korrelation der Variablen mit der Erste Hauptkomponente beschrieben. Die Betrachtung der LKR-LCL zeigt uns, dass drei Cluster sich bilden lassen(Bild 3.15). Eine Negative Korrelation zwischen der blauen Variablen und der grauen Variablen lässt

sich hier bilden. Das heißt, wenn die blauen Variablen zunehmen, die grauen abnehmen und umgekehrt. Ferner es ist auffällig, dass die gelben Variablen unabhängig, sowohl von der blauen als auch von der grauen sind. Die Werte für die LKL-LCR befinden sich im Anhang.

	correlation	p.value
xVelocity	0.79047548	0.000000e+00
maneuver	0.64628165	0.000000e+00
class	0.62018449	0.000000e+00
timeGapFolloAdj	0.52430960	0.000000e+00
distFolloAdj	0.52036455	0.000000e+00
distFollo	0.40453538	2.451316e-191
timeGapFollo	0.33193527	1.084284e-125
classFollo	0.22607997	1.576233e-57
distPredAdj	0.21772054	2.198263e-53
classPred	0.17671470	1.752829e-35
timeGapPredAdj	0.09474262	3.412342e-11
xAcceleration	0.04472538	1.788832e-03
distPred	0.03362295	1.890414e-02
accFolloAdj	-0.05046564	4.241993e-04
accFollo	-0.07979532	2.428789e-08
timeGapPred	-0.13259779	1.459863e-20
accPredAdj	-0.15883055	6.645109e-29
diffSpeedFollo	-0.36337735	4.771358e-152
yAcceleration	-0.48113435	6.567199e-281
yVelocity	-0.54029444	0.000000e+00
diffSpeedFolloAdj	-0.56609884	0.000000e+00
diffSpeedPredAdj	-0.57326748	0.000000e+00
diffSpeedPred	-0.62591752	0.000000e+00

Bild 3.15: Korrelation zwischen Variablen und Erste Hauptkomponente

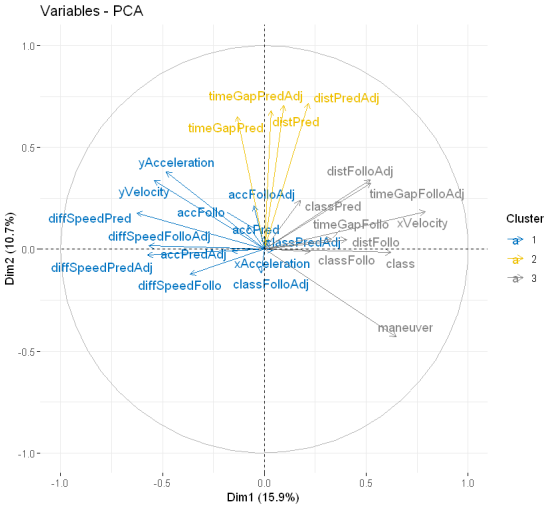


Bild 3.16: Korrelationskreis

3.1.4 Zusammenfassung der Beschreibende Statistik

Die PCA ermöglicht uns eine Reduktion der Variablen von 25 auf 16 Variablen und damit einen besseren Umgang mit den Daten. Die Berechnung der Eigenvalue und die Verteilung der Information in den Komponenten hilft zwar nicht, eine aussagekräftige Beurteilung zu treffen, wenn man die Zwei Dimensionale Betrachtungen der Variablen auf die Erste und ZweiteKomponenten beobachtet. Da die beiden Komponenten insgesamt 26,6% Varianz Anteil haben, für den Fall LCL-LKR und 23,5% für den Fall LKL-LCR. Aber wir können einige Verhältnisse zwischen den Variablen beobachtet. Laut PCA wissen wir, dass einige Variabel keine Rolle spielen oder manche ganz wenig für eine bestimmte Hauptkomponente. Es wird für die vorhersagende Analyse vermutet, ob ein Fahrzeug einen Spurwechsel unternimmt oder sein Spur bei behalten will. Es gibt einige Variablen, die gar keine Einflüsse auf die Manöver haben. Es wird anhand der Entscheidungsbaum Methode versucht, die Variablen, bezüglich der Manöver zu klassieren nach ihrer Wichtigkeit.

3.2 Decision tree Analyse zur Auswahl der erklärenden Variablen

Die Entscheidungsbaummethode ist ein statistisches Werkzeug zur Klassifizierung, Vorhersagen, Interpretation und Datenmanipulation, und wird oft in der Forschung verwendet. Die Ergebnisse der Entscheidungsbaum Methoden hat 9 Variablen für den Fall LKR-LCL ergeben, die für die Lane changing auf der linken Spur oder der Spur Beibehaltung eine entscheidende Rolle spielen.(siehe Bild 3.16). Andererseits hat die Methode, für den Fall LKL-LCR 5 Variablen ergeben die entscheidend sind, ob ein Fahrzeug die rechte Spur nehmen soll oder sein Spur lieber beibehält. Ein großer Vorteil lässt sich hier ableiten. 9 Variablen bzw. 5 Variablen werden weiter betrachtet und eine Prädiktion mit Logistischen Regression gemacht.

	index	variabel	Wichtigkeit
0	1	yVelocity	0.818598
1	3	yAcceleration	0.147621
2	21	diffSpeedFolloAdj	0.013211
3	23	accFolloAdj	0.004268
4	10	distFollo	0.004066
5	2	xAcceleration	0.003910
6	6	diffSpeedPred	0.003060
7	7	timeGapPred	0.002478
8	12	timeGapFollo	0.001395
9	14	classFollo	0.001395
10	16	diffSpeedPredAdj	0.000000
11	22	timeGapFolloAdj	0.000000
12	20	distFolloAdj	0.000000
13	19	classPredAdj	0.000000
14	18	accPredAdj	0.000000
15	17	timeGapPredAdj	0.000000
16	0	xVelocity	0.000000
17	15	distPredAdj	0.000000
18	13	accFollo	0.000000
19	11	diffSpeedFollo	0.000000
20	9	classPred	0.000000
21	8	accPred	0.000000
22	5	distPred	0.000000
23	4	class	0.000000
24	24	classFolloAdj	0.000000

Bild 3.17: Ergebnis der Entscheidungsbaum LKR-LCL

	index	variable	Importance
0	1	yVelocity	0.988228
1	2	xAcceleration	0.004874
2	16	diffSpeedPredAdj	0.003286
3	23	accFolloAdj	0.001630
4	8	accPred	0.001310
5	10	distFollo	0.000672
6	0	xVelocity	0.000000
7	15	distPredAdj	0.000000
8	22	timeGapFolloAdj	0.000000
9	21	diffSpeedFolloAdj	0.000000
10	20	distFolloAdj	0.000000
11	19	classPredAdj	0.000000
12	18	accPredAdj	0.000000
13	17	timeGapPredAdj	0.000000
14	12	timeGapFollo	0.000000
15	14	classFollo	0.000000
16	13	accFollo	0.000000
17	11	diffSpeedFollo	0.000000
18	9	classPred	0.000000
19	7	timeGapPred	0.000000
20	6	diffSpeedPred	0.000000
21	5	distPred	0.000000
22	4	class	0.000000
23	3	yAcceleration	0.000000
24	24	classFolloAdj	0.000000

Bild 3.18: Ergebnis der Entscheidungsbaum LKl-LCR

3.3 Vorhersagen von Spurwechsel Anhand Logistische Regression

3.3.1 Train-, Test-, Validation Datensätze

Um das Modell zu trainieren, wird zuerst das entsprechende Modell aus scikitlearn importieren, dann wird die Anpassungsmethode sklearn .model_selection.train_test_split aufgerufen um die Daten in 6 Teile zu teilen.

Tabelle 3.1: Daten Teilung

x_train	x_test	x_validation	y_train	y_test	y_validation
(3509, 9)	(975, 9)	(390, 9)	(3509,)	(975,)	(390,)

3.3.2 Ergebnis

Die Modelle liefert in Bild 3.18 und 3.19 die Regression Koeffizient an.

	variabel	Regressionskoefizient
0	[xVelocity]	[-8.049239502025861]
1	[yVelocity]	[-5.635426025009274]
2	[xAcceleration]	[-0.08331992159766456]
3	[yAcceleration]	[-0.6824203020010048]
4	[class]	[0.013411781820436914]
5	[distPred]	[1.098960887634679]
6	[diffSpeedPred]	[-0.4033859262864567]
7	[timeGapPred]	[-1.0809300821345669]
8	[accPred]	[-0.3618620463474219]

Bild 3.19: Ergebnis der logistische Regression Analyse LKR-LCL

	variabel	Regressionskoefizient
0	[xVelocity]	[-0.4808627476222996]
1	[yVelocity]	[-0.8812185842410054]
2	[xAcceleration]	[0.31498960660815684]
3	[yAcceleration]	[-0.2430399403908782]
4	[class]	[-0.040613601031735454]

..

Bild 3.20: Ergebnis der logistische Regression Analyse LKL-LCR

$$P_{LKR-LCL} = \frac{e^{(a_0+a_1x_1+a_2x_2+a_3x_3+a_4x_4+a_5x_5+a_6x_6+a_7x_7+a_8x_8)}}{1+e^{(a_0+a_1x_1+a_2x_2+a_3x_3+a_4x_4+a_5x_5+a_6x_6+a_7x_7+a_8x_8)}} \tag{3.1}$$

3.3.3 Evaluation

Präzision ist die Fähigkeit des Modells, die tatsächlich nicht Spurwechseln als Spurwechsel zu kennzeichnen. Für jede Klasse ist es definiert als das Verhältnis von True Positives zur Summe von True Positive und False Positive. Für den Fall LKR-LCL die Präzision beträgt 99% und für LKL-LCR 92%. Recall ist die Fähigkeit unserer Modelle, alle tatsächliche Spurwechsel zu finden und das Verhältnis von True Positiven zu der Summe von True Positiven und fasle Negativen und beträgt 99% für LKR-LCL und 91% LKL-LCR. .Der „F-Score“ ist ein Maß für die „Gesamtgüte“ und beträgt 99% für LKR-LCL und 88% LKL-LCR.

	precision	recall	f1-score	support
1	0.99	1.00	1.00	3229
4	0.99	0.94	0.96	280
avg / total	0.99	0.99	0.99	3509

Bild 3.21: Evaluation LKR-LCL

	precision	recall	f1-score	support
2	0.92	0.99	0.95	4465
3	0.57	0.10	0.17	446
avg / total	0.89	0.91	0.88	4911

Bild 3.22: Evaluation LKL-LCR

3.4 Netlogo

Netlogo ist eine OpenOffice Softwarepaket und eine programmierbare Modellierungsumgebung mit mehreren Agenten. Es wird von vielen hunderttausenden Schülern, Lehrern und Forschern weltweit verwendet. Es wurde von Uri Wilensky verfasst und am (The Center for Connected Learning and Computer-Based Modeling) der Northwestern University entwickelt. [32]

3.4.1 Oberfläche und Simulationsparameter

Das Bild 3.22 zeigt die Interface von Netlogo. Es besteht hauptsächlich aus 7 Tasten und einem Ausgebefenster. Ein Klick aus das Setup, lässt die Fahrbahn mit zwei Fahrstreife auf die Ausgeberfenster (Bild 3.23) zeichnen. Die On-Off Tasten dient zur Auszeichnung der Trajektorien während der Fahrt. Der einstellbare Regressionskoffizient für die linke Spur (a_{0L},...a_{8L}) und für die rechte Spur (a_{0L},...a_{8L} .Ein Klick auf der Set-ai stellt sich die Regressionskoefizinten auf das kalibrierte Modell. Die move-Tasten erlauben uns die Bewegung der Fahrzeuge. Die Bewegung der Fahrzeuge, entspricht der Realität, weil das Implementieren beruht auf den HighD Messdaten.

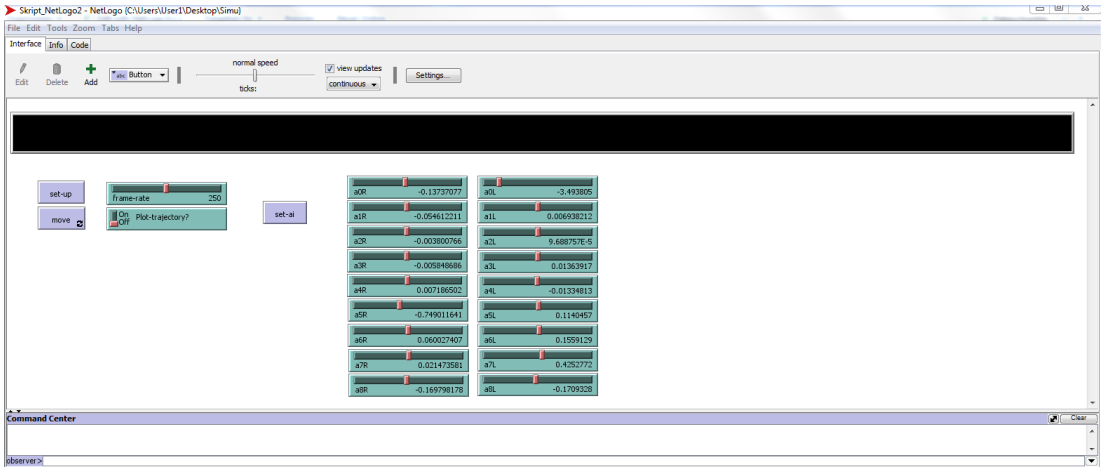


Bild 3.23: Simulationsoberfläche

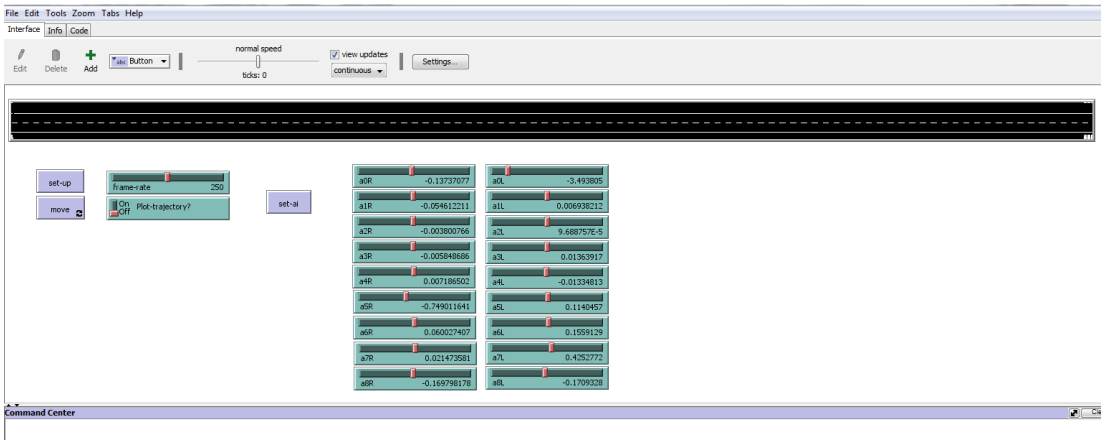


Bild 3.24: Simulationsoberfläche

3.4.2 Durchführung

Der Klick auf die move Tasten setzen die Fahrzeuge in Bewegung. (Bild 3.25) Jedes Fahrzeug lässt sich mit seine Id, xVellocity, yVellocity , dispred, distFollo, diffSpeedPred. Die Fahrzeuge, die vorne kurzen Abstand haben, sind mit Hell-Rot gekennzeichnet (Bild 3.25) und die hinten stehen, haben die Farbe Grün. Es wird während der Fahrt kontinuierlich die P Werte für jedes Fahrzeug gemessen und wenn der P-Wert gegen 1 läuft, wird die grüne Farbe zu dunkelrot, das bedeutet, dass es die logistische Regression vorhersagt , dass das entsprechende Auto die Spur wechseln soll. Ein P-Wert, der gegen 0 läuft, bedeutet Lane-Keeping. Der Befehl für jedes Auto lautet: Gibt es andere Autos vor dir?, wenn ja misst seine Geschwindigkeit , sein Abstand, Geschwindigkeitsunterschied ,usw. und genauso wenn ein anderes Auto hinten ist.

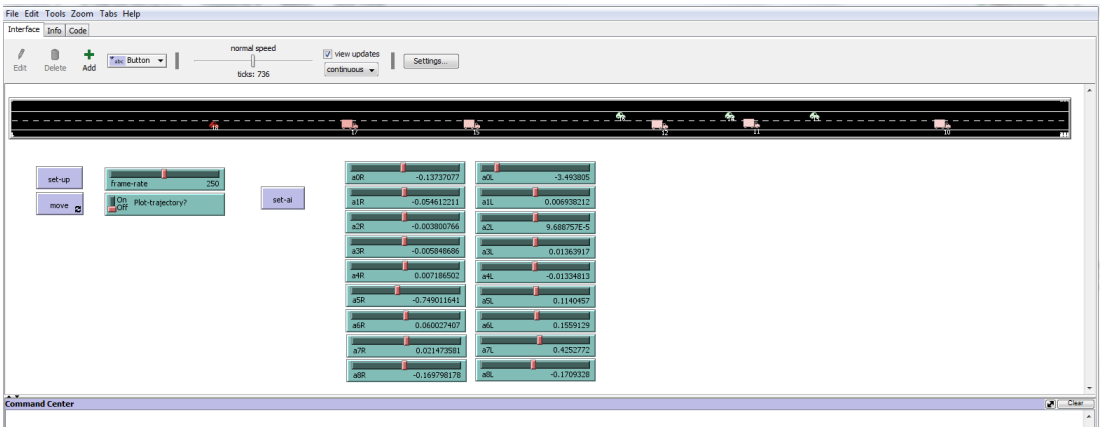


Bild 3.25: Simulationsoberfläche

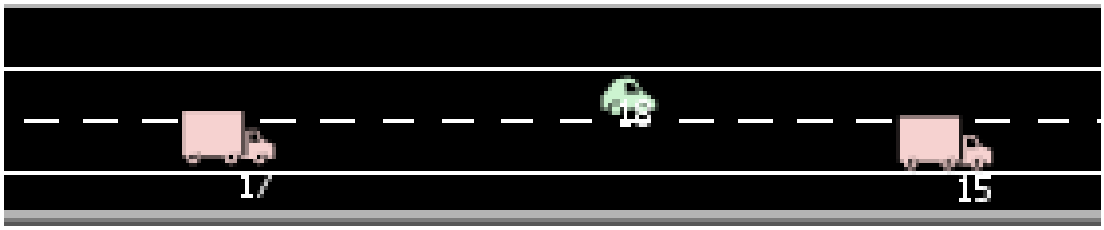


Bild 3.26: Simulationsoberfläche

3.4.3 Ergebnis und Interpretation

Qualitative

Folgender Sachverhalt lässt sich fest stellen mit der Simulation: Es gibt viele Fälle wo das Modell eine Lane changing vorhersagt und es erfolgt tatsächlich eine Lane changing(763,181). Beispiel Fahrzeug 18. Es gibt wiederum Fälle wo das Modell keine lane changing vorhersagt, aber es erfolgt eine lane changing (True Negative-Fehler) Beispiel (Fahrzeuge 456,1088). Es gibt Fälle wo das Modell zwar eine Lane-changing zu früh Vorhersagt aber es erfolgt keine Lane Changing(946). (False-Positiv-Fehler).

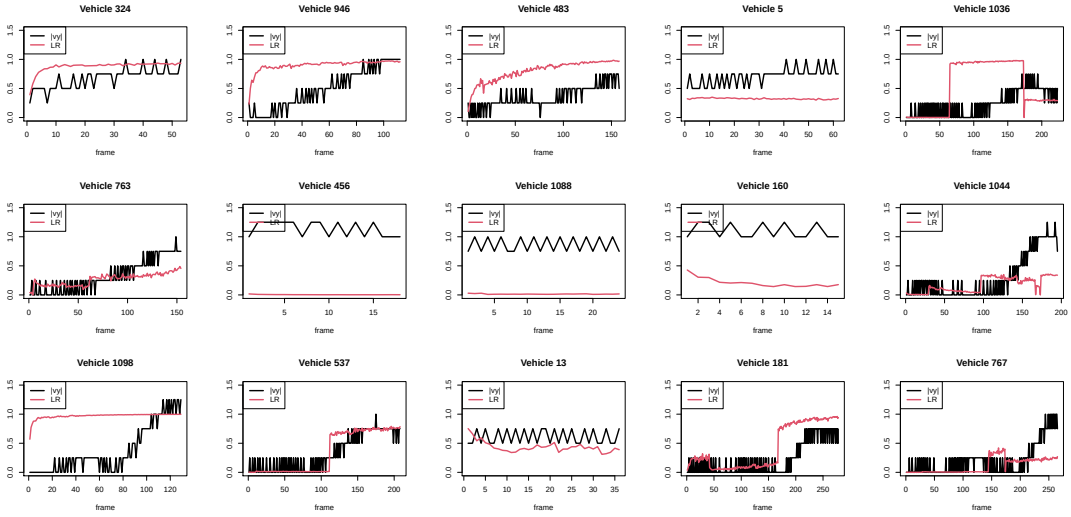


Bild 3.27: Simulationsergebnis

Quantitative

An dieser Stelle wird es untersucht ob der P_Werte mit der yVelocity korreliert. Da während der lane changing wenn der P_Werte steigt, normalerweise auch die Geschwindigkeit im y-Richtung auch zunehmen muss. Für die Analyse benötigen wir einige Daten. Die Ergebnisse anhand der Pearson Korrelation gibt uns eine schwache Korrelation im Werte von 0,32 .

4 Schluss

Die deskriptive Analyse hilft zwar, die Beziehung zwischen den Variablen miteinander zu verstehen aber aufgrund der Vielzahl der Variablen ist eine Ansage über einen möglichen Trend zu Treffen oder irgend welche Phänomene in den Datensätze zu erkennen. Die Methoden der Hauptkomponente Analyse, ermöglichen uns zwar keine signifikante Beschreibung(2D) der Daten über die ersten und zweiten Komponenten, da die ersten beiden Komponenten eine Anteil kleiner als 30% sind, aber es gibt uns einige Informationen über die Daten in 2 Dimensionalräumen. Beispielsweise 3 Muster in den Daten lässt sich erkennen durch die Darstellung in der Korrelation zirkle. Das Ergebnis der Prädiktion durch die Simulation mit der logistische Regression hat in einige Fällen gut funktioniert und in Anderen nicht so gut. Wir gehen davon aus, dass die negativen Ergebnisse einfach an dem Faktor Menschen liegen. Einfache Beispiele: Es kommt vor, dass das Fahrzeug vorne eine kleinere Geschwindigkeit hat, aber anstatt es zu überholen, wird versucht die eigene Geschwindigkeit zu reduzieren, weil die nächste Ausfahrt beispielsweise nahe ist. Wenn es um ein autonomes Fahrzeug geht, wäre das Ergebnis anders, da die Fahrzeuge, denselben Regeln folgen würden.

Literaturverzeichnis

[1] Statistische Bundesamt (Destatis). Alle 9 Stunden stirbt ein Mensch bei einem Geschwindigkeitsunfall in Deutschland; 2020. Abgerufen: 03.02.2021. Available from: https://www.destatis.de/DE/Presse/Pressemitteilungen/2020/07/PD20_N039_46241.html.

[2] H Jula E, P A Ioannou. Collision avoidance-analysisfor lane changing and merging; 2000. Abgerufen: 03.02.2021. Available from: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=901899.

[3] Petros A Ioannou HE;. .

[4] R DeSonia DGPG. CAPC: A Road-Departure Prevention System; 2000. Abgerufen: 04.02.2021. Available from: https://ieeexplore.ieee.org/abstract/document/546271.

[5] C R Kelber C. A lane departure warning system based on a linear-parabolic lane model; 2004. Abgerufen: 04.02.2021. Available from: https://ieeexplore.ieee.org/abstract/document/1336503.

[6] volkswagen. Fahrerassistenzsysteme; 2021. Abgerufen: 04.02.2021. Available from: https://www.volkswagen-nutzfahrzeuge.de/de/alles-ueber-volkswagen-nutzfahrzeuge/fahrerassistenz-und-sicherheitssysteme.html.

[7] Christian Kellner). SicherheitSchriftenreihe Verkehrssicherheit; 2021. Abgerufen: 04.02.2021. Available from: https://www.dvr.de/fileadmin/downloads/dvr-schriftenreihe/Schriftenreihe-Verkehrssicherheit-13.pdf.

[8] Gabriele Wolf (Hrsg) HS. Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort. In: Winner H, editor. Handbuch Fahrerassistenz-systeme. ieweg +Teubner Verlag: 2., korrigierte Auflage; 2012. p. 15. Available from: ISBN 978-3-8348-1457-9.

[9] Philipp Ring. Modellierung kognitiver Prozesse für ein numerisches Fahrerverhaltensmodell. In: VERLAG C, editor. Handbuch Fahrerassistenz-systeme. eISBN 978-3-7369-6238-5: 2., korrigierte Auflage; 2020. p. 11. Available from: eISBN 978-3-7369-6238-5.

[10] Marc Semrau. Untersuchung zur Modellierung von chinesischem Fahrverhalten aufAutobahnen für den Test pilotierter Fahrfunktionen. In: Aktiengesellschaft V, editor. Untersuchung zur Modellierung von chinesischem Fahrverhalten aufAutobahnen für den Test pilotierter Fahrfunktionen. Springer,Band 130; 2018. p. 11. Available from: https://doi.org/10.1007/978-658-23761-5.

[11] Gabriele Wolf (Hrsg) HS. Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort. In: Winner H, editor. Handbuch Fahrerassistenz-systeme. ieweg +Teubner Verlag: 2., korrigierte Auflage; 2012. p. 16. Available from: ISBN 978-3-8348-1457-9.

[12] Reichart, G B. Braess, H.-H.; Reichart, G. (1995). PROMETHEUS - Vision des intelligenten Automobi1s auf intelligenter StraBe; 1995. Abgerufen: 12.02.2021. Available from: ATZ 97, Heft 4+6.

[13] Klaus-Peter Timpe (Hrsg) T. In: Kraftfahrzeugführung. Springer; 2001. p. 31–32. Abgerufen: 12.02.2021. Available from: ISBN 978-3-642-56721-6 (eBook).

[14] Konrad Reif (Hrsg). In: Winner H, editor. Fahrstabilisierungssysteme und Fahrerassistenzsysteme. 1. Auflage 2010,Vieweg+Teubner Verlag ,Springer: 2., korrigierte Auflage; 2010. p. 104. Available from: ISBN 978-3-8348-1314-5.

[15] Lerner, M ; Albrecht, M ; Evers, C. Das Unfallgeschehen bei Nacht, Bericht der Bundesanstalt für Straßenwesen (BASt), Heft M 172; 2005. Accessed: 24.06.2020. Available from: https://bast.opus.hbz-nrw.de/opus45-bast/frontdoor/deliver/index/docId/1735/file/M172.pdf.

[16] Wolfgang Huhn T. Sichtverbesserungssysteme. In: Singer C, editor. Handbuch Fahrerassistenzsysteme,Grundlagen, Komponenten und Systeme fur aktive Sicherheit und Komfort. Springer Fachmedien Wiesbaden: 3., uberarbeitete und erganzte Auflage; 2015. p. 817. Available from: ISBN 978-3-658-05734-3.

[17] Gabriele Wolf (Hrsg) H. In: Winner H, editor. Handbuch Fahrerassistenzsysteme. Vieweg+Teubner Verlag |Springer Fachmedien Wiesbaden GmbH 2012: 2., korrigierte Auflage; 2012. p. 448–461. Available from: ISBN 978-3-8348-1457-9.

[18] Gabriele Wolf (Hrsg) H. In: Winner H, editor. Handbuch Fahrerassistenzsysteme. Vieweg+Teubner Verlag |Springer Fachmedien Wiesbaden GmbH 2012: 2., korrigierte Auflage; 2012. p. 548–551. Available from: ISBN 978-3-8348-1457-9.

[19] Li Liu Yongsheng Ou G, Zhangjun Song. Dynamic Modeling of Driver Control Strategy of Lane-Change Behavior and Trajectory Planning for Collision Prediction; 2012. Available from: DOI: 10.1109/TITS.2012.2187447.

[20] P G Gipps,. “A model for the structure of lane-changing decisions,”; 1986. Available from: Transp. Res. B: Methodol., vol. 20, no. 5, pp. 403\T1\textendash 414,.

[21] J C McCall, D P Wipf, M M Trivedi, and B D Rao. “Lane change intent analysis using robust operators and sparse Bayesian learning,”. berlin; 2007. Available from: IEEE Trans. Intell. Transp. Syst., vol. 8, no. 3, pp. 431\T1\textendash 440,.

[22] A Pentland and A Liu,. “Modeling and prediction of human behavior,” Neural Comput., vol. 11, no. 1, pp. 229–242,; 1999. Available from: Neural Comput., vol. 11, no. 1, pp. 229\T1\textendash 242.

[23] D D Salvucci,. “Modeling driver behavior in a cognitive architecture,” Summer 2006.; 2019. Available from: Hum. Factors, vol. 48, no. 2, pp. 362\T1\textendash 380,.

[24] J H Kim, S Hayakawa, T Suzuki, K Hayashi, S Okuma, N Tsuchida, M Shimizu, and S Kido,. “Modeling of driver’s collision avoidance maneuver based on controller switching model,”; 2005. Available from: IEEE Trans. Syst., Man, Cybern. A: Syst., Humans, vol. 35, no. 6, pp. 1131\T1\textendash 1143,.

[25] Kazi Iftekhar Ahmed. Modelling Drivers‘Accelaration and lane changing Behaviour; 1999. .

[26] matchingHartung Elpert. Multivariate Statik, Lehr und Handbuch der angewandten Statistik,6.Auflage; 1999. Available from: ISBN 3-486-25287-9.

[27] Robert Krajewski LL Julian Bock. The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. EEE 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, Hawaii, USA; 2018. Available from: https://arxiv.org/ftp/arxiv/papers/1810/1810.05642.pdf.

[28] Basma Khelfa1and Antoine Tordeux, Antoine Tordeux . Understanding and predicting lane-changing intends on highwaysusing logistic regressions; 2021. .

[29] A Puri . A survey of unmanned aerial vehicles (UAV) for traffic surveillance; 2005. Available from: University of South Florida, Florida.

[30] Puri K, M Kontitsis . Statistical profile generation for traffic monitoring using real-time UAV based video data; 2007. Available from: Mediterranean Conference on Control & Automation, Athens.

[31] Herve Abdi and Lynne J Williams tPuh year = 2010;. .

[32] ;. Accessed: 19.06.2020. Available from: https://ccl.northwestern.edu/netlogo/.

Abkürzungsverzeichnis

highD highway drone dataset.
AuG Abstands- und Geschwindigkeitsregelanlage
xVelocity Geschwindigkeit des Fahrzeugs
yVelocity Seitliche Geschwindigkeit des Fahrzeugs
xAcceleration Beschleunigung des Fahrzeugs
yAcceleration Seitliche Beschleunigung des Fahrzeugs
distPred Abstand zwischen Fahrzeug
diffSpeedPred Geschwindigkeitsdifferenz zwischen Fahrzeug
timeGapPred Zeitlicher Abstand zwischen Fahrzeug
accPredBeschleunigung des Fahrzeugs
classPred Fahrzeugart (PKW oder LKW) des Fahrzeugs
distFollo Abstand zwischen Fahrzeug
diffSpeedFollo Geschwindigkeitsdifferenz zwischen Fahrzeug
timeGapFollo Zeitlicher Abstand zwischen Fahrzeug
accFollo Beschleunigung des Fahrzeugs
classFollo Fahrzeugart (PKW oder LKW) des Fahrzeugs
distPredAdj Abstand zwischen Fahrzeug
diffSpeedPredAdj Geschwindigkeitsdifferenz zwischen Fahrzeug
timeGapPredAdj Zeitlicher Abstand zwischen Fahrzeug
accPredAdj Beschleunigung des Fahrzeugs
classPredAdj Fahrzeugart (PKW oder LKW) des Fahrzeugs
distFolloAdj Abstand zwischen Fahrzeug
diffSpeedFolloAdj Geschwindigkeitsdifferenz zwischen Fahrzeug
timeGapFolloAdj Zeitlicher Abstand zwischen Fahrzeug
accFolloAdj Beschleunigung des Fahrzeugs
classFolloAdj Fahrzeugart (PKW oder LKW) des Fahrzeugs
diffaccPred Beschleunigungsdifferenz zwischen Fahrzeug
diffaccFollo Beschleunigungsdifferenz zwischen Fahrzeug
diffaccPredAd Beschleunigungsdifferenz zwischen Fahrzeug
diffaccPredFollo Beschleunigungsdifferenz zwischen Fahrzeug

Bilderverzeichnis

2.1 Drei-Ebenen-Modell nach Rasmussen [8] 4
2.2 Drei-Ebenen-Modell nach Donges [8] 5
2.3 Fahrassistenzsysteme im Zusammenhang mit Fahr-aufgaben [13] 6
2.4 Häufigkeitsverteilung von Nachtunfall [16] 7
2.5 Nachtunfall Ursachen [16] 8
2.6 Zusammenhang zwischen Bremsweg und Fahrgeschwindigkeit [16] 10
2.7 Markierungslicht [16] 11
2.8 Assistenzstrategien in kritischen Verkehrssituationen [17] 12
2.9 Struktureller Aufbau von Spurwechsel [25] 17
2.10 Darstellung von Fahrbahn [16] 19
2.11 Schematische Modellierung von Regression Logstik [16] 22

3.1 Drohnen Messung [28] 25
3.2 Drohnen Messung 25
3.3 Paarweise Darstellung der Variable LCL-LKR 26
3.4 Paarweise Darstellung der Variabel LKL-LCR 26
3.5 Lage Parameter 27
3.6 Lage Parameter 27
3.7 Korrelation Tabelle LCL-LKR 28
3.8 Korrelation Tabelle LCR-LKR 28
3.9 Prinzipal Komponente LCL-LKR 29
3.10 Prinzipal Komponente LKL-LCR 29
3.11 Beitrag der Variabel LCL-LKR 30
3.12 Beitrag der Variabel LKL-LCR 30
3.13 Korrelationskreis LCL-LKR 31
3.14 Korrelationskreis LKL-LCR 31
3.15 Korrelation zwischen Variablen und Erste Hauptkomponente 32
3.16 Korrelationskreis 32
3.17 Ergebnis der Entscheidungsbaum LKR-LCL 34
3.18 Ergebnis der Entscheidungsbaum LKl-LCR 34
3.19 Ergebnis der logistische Regression Analyse LKR-LCL 35
3.20 Ergebnis der logistische Regression Analyse LKL-LCR 35
3.21 Evaluation LKR-LCL 36
3.22 Evaluation LKL-LCR 36
3.23 Simulationsoberfläche 38
3.24 Simulationsoberfläche 38
3.25 Simulationsoberfläche 39
3.26 Simulationsoberfläche 39
3.27 Simulationsergebnis 40

Tabellenverzeichnis

2.1 Eigenschaften der Lichtquelle eines Frontscheinwerfers 9
2.2 Sichtbarkeitsweite Lichtquelle eines Frontscheinwerfers 9
2.3 Spurwechselnde Modelle [19] 16

3.1 Daten Teilung 35

Eidesstattliche Erklärung

- *Hiermit, erkläre ich, dass ich die von mir eingereichte Abschlussarbeit (Bachelor-Thesis) selbstständig verfasst und keine andere als die angegebene Quelle und Hilfsmittel benutzt sowie Stellen der Abschlussarbeit, die anderen Werken dem Wortlaut oder Sinn nach entnommen wurden, in jedem Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.*

  Ich bin damit einverstanden, dass die Arbeit durch Dritte eingesehen und unter Wahrung urheberrechtlicher Grundsätze zitiert werden darf.

Ort und Datum: _____

Unterschrift: _____

Anhang

25_variable

March 29, 2021

Python_Skript

0.0.1 LKL und LCR

0.1 Datenmanipulation

raw_data= pd.read_csv("/Users/mlg7777/Desktop/data_finala1/LKR_LCL.csv")
raw_data.head()
raw_data.columns
raw_data=raw_data.drop(["X","Unnamed: 0"],axis=1)
for column in raw_data: unik_vals=np.unique(raw_data[column])

```
nr_value=len(unik_vals)
if nr_value <10:
    print("Anzahl der Elemente {}  :{}--{}".format(column,nr_value,unik_vals))
else: print("Anzahl der Elemente {}  :{}".format(column,nr_value))
```

LCL=maneuver==4 lane changing to the left lane (right->left)

LKL=maneuver==2 #lane keeping to the right lelft

LCR=maneuver==3 lane changing to the right lane (left->right)

LKR=maneuver==1 #lane keeping to the right lane

0.2 Explorative Datenanalyse

```
In [144]: raw_data.describe()
```

```
max       1.150000       1.000000

[8 rows x 26 columns]
```

```
In [145]: raw_data.columns
```

```
Out[145]: Index(['maneuver', 'xVelocity', 'yVelocity', 'xAcceleration', 'yAcceleration',
                 'class', 'distPred', 'diffSpeedPred', 'timeGapPred', 'accPred',
                 'classPred', 'distFollo', 'diffSpeedFollo', 'timeGapFollo', 'accFollo',
                 'classFollo', 'distPredAdj', 'diffSpeedPredAdj', 'timeGapPredAdj',
                 'accPredAdj', 'classPredAdj', 'distFolloAdj', 'diffSpeedFolloAdj',
                 'timeGapFolloAdj', 'accFolloAdj', 'classFolloAdj'],
                dtype='object')
```

```
In [146]: raw_data.to_csv("/Users/mlg7777/Desktop/data_finala1/LKR_LCL1.csv")
```

```
In [147]: from sklearn.preprocessing import MinMaxScaler
          import pandas as pd
          from sklearn.preprocessing import StandardScaler
          #scaler=MinMaxScaler()
          scaler=StandardScaler()
          df=scaler.fit(raw_data)
          df
```

```
Out[147]: StandardScaler(copy=True, with_mean=True, with_std=True)
```

scaled_data=scaler.transform(raw_data) scaled_data= pd.DataFrame(scaled_data) scaled_data

```
In [149]: scaled_data.columns=['maneuver', 'xVelocity', 'yVelocity', 'xAcceleration',
                  'yAcceleration', 'class', 'distPred', 'diffSpeedPred', 'timeGapPred',
                  'accPred', 'classPred', 'distFollo', 'diffSpeedFollo', 'timeGapFollo',
                  'accFollo', 'classFollo', 'distPredAdj', 'diffSpeedPredAdj',
                  'timeGapPredAdj', 'accPredAdj', 'classPredAdj', 'distFolloAdj',
                  'diffSpeedFolloAdj', 'timeGapFolloAdj', 'accFolloAdj', 'classFolloAdj']
```

scaled_data

0.3 Korrelation

```
In [151]: pearsonr_coefficient ,p_value = pearsonr(scaled_data['distPred'],scaled_data['xVelocity'])
          print("pearsonR correlation coeffizient %0.3f"%(pearsonr_coefficient))
```

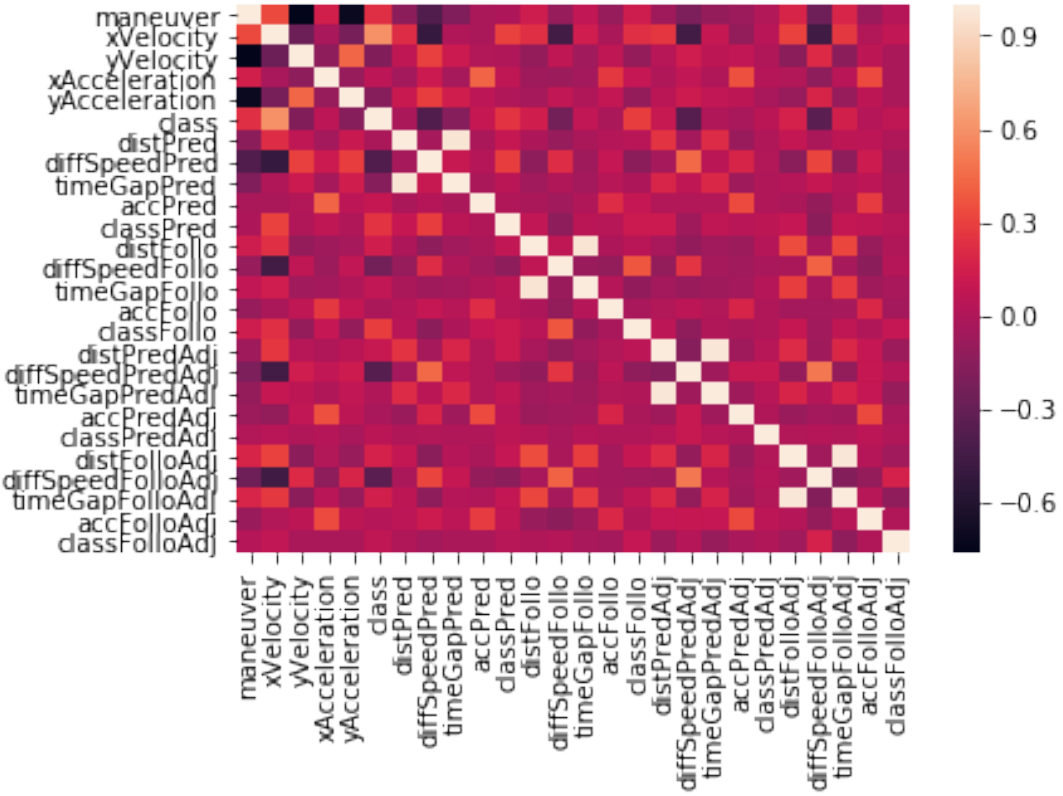
```
pearsonR correlation coeffizient 0.220
```

```
In [152]: correlation_Matrix =scaled_data.corr()
```

correlation_Matrix
for column in correlation_Matrix: print("positive correlation :{}--{}".format(correlation_Matrix[column]>0.1,correlation_Matrix[column]))

In [155]: sns.heatmap(correlation_Matrix, xticklabels=correlation_Matrix.columns.values,

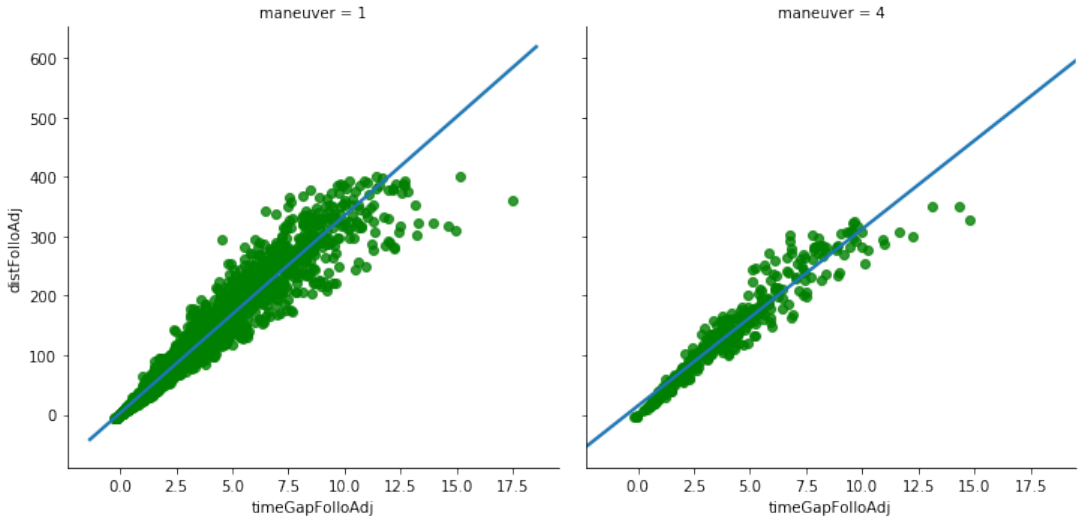
yticklabels=correlation_Matrix Out[155]: <matplotlib.axes._subplots.AxesSubplot at 0x1a1de06588>

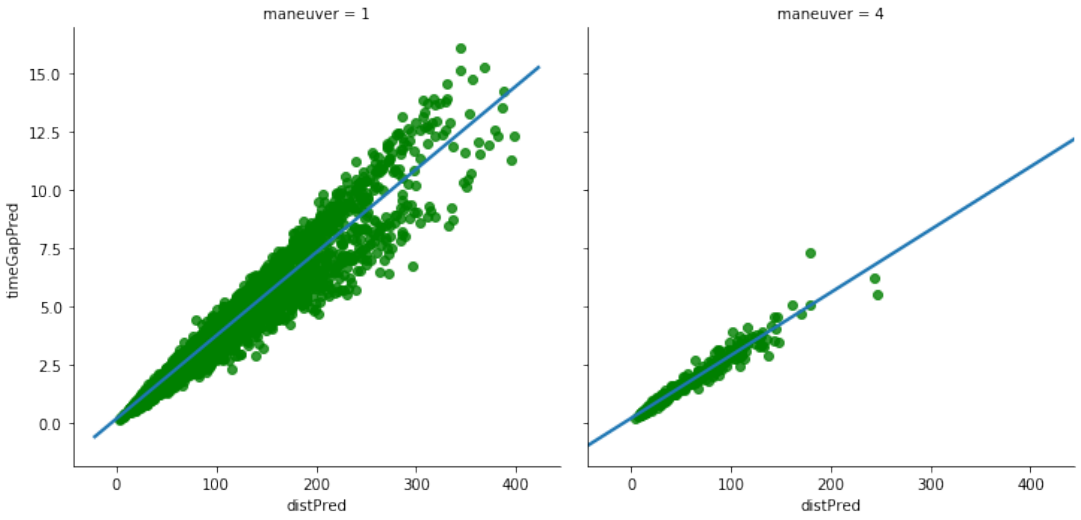


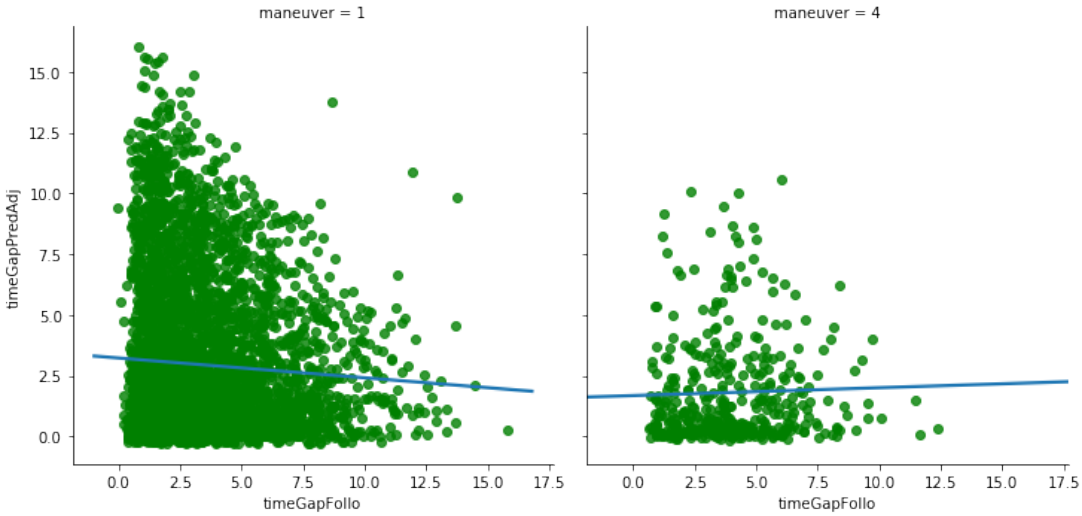
In [ ]:

In [156]: g=sns.lmplot(x='timeGapFolloAdj',y='distFolloAdj',data=raw_data,col="maneuver",col_wrap=

```
a=sns.lmplot(x='distPred',y='timeGapPred',data=raw_data,col="maneuver",col_wrap=3,height=5,scatter_kws
b=sns.lmplot(x= 'timeGapFollo',y='accPred',data=raw_data,col="maneuver",col_wrap=3,height=5
c=sns.lmplot(x= 'classFollo',y='distPredAdj',data=raw_data,col="maneuver",col_wrap=3,height
d=sns.lmplot(x= 'diffSpeedPredAdj',y='accFolloAdj',data=raw_data,col="maneuver",col_wrap=3,height
e=sns.lmplot(x= 'distFolloAdj',y='classPredAdj',data=raw_data,col="maneuver",col_wrap=3,height
f=sns.lmplot(x= 'timeGapFollo',y='timeGapPredAdj',data=raw_data,col="maneuver",col_wrap=3,height
  /anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using a non-tuple
            return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```


maneuver = 1
maneuver = 4
distFolloAdj
timeGapFolloAdj
0
100
200
300
400
500
600
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


maneuver = 1
maneuver = 4
timeGapPred
distPred
0.0
2.5
5.0
7.5
10.0
12.5
15.0
0
100
200
300
400



import numpy as np import pandas as pd import matplotlib.pyplot as plt from sympy import * import math

import scipy.integrate as integrate import scipy.special as special re= integrate.quad(lambda x:special.jv(2.5,x),0,4.5) re

vit=LCR_data["xVelocity"] vit

```
In [ ]: def mean(array):
            n=len(array)
            return 1/n*sum(array)
        def varianz(array):
            n=len(array)
            return (1/(n-1))*sum(map(lambda Xi:(Xi-ţ)**2,array))
```

```
In [ ]: x=raw_data[["xVelocity","xAcceleration",'distPred','diffSpeedPred','timeGapPred']]
        xvelo=raw_data["xVelocity"]
        xaccel=raw_data["xAcceleration"]
        distpre=raw_data["distPred"]
        diffspepre=raw_data["diffSpeedPred"]
        timegappre=raw_data["timeGapPred"]
        pearsonr_coefficient,p_value=pearsonr(xvelo,distpre)
        print("pearsonR correlation coefficient %0.3f" % (pearsonr_coefficient))
```

```
In [ ]: corr= x.corr()
```

```
In [ ]: corrglob= raw_data.corr()
```

```
In [ ]: import scipy
        from pylab import rcParams
        import seaborn as sb
        from scipy.stats.stats import pearsonr
        import numpy as np
```

```
        import pandas as pd
        import matplotlib.pyplot as plt
        from sympy import *
        import math
        from scipy.integrate import *
        from scipy.interpolate import*
```

In [ ]:
```
        import matplotlib.pyplot as plt
        import numpy as np
        import datetime
        import matplotlib.dates as mdates
        from matplotlib.ticker import AutoMinorLocator


        diffspepre=np.array(raw_data["diffSpeedPred"])
        xvelo=np.array(raw_data["xVelocity"])
        x_p = np.linspace(-25,5,1000)
        p3 = np.polyfit(diffspepre,xvelo,3)
        y= np.polyval(p3,x)
        plt.plot(diffspepre, xvelo, "o")
        plt.plot(x_p, y_p,"-")
        plt.legend(["data","polyfit"],loc="best")
        #plt.ylim(-2,2)
```

In [ ]:
```
        diffspepre=np.array(raw_data['timeGapFolloAdj'])
        xvelo=np.array(raw_data["distFolloAdj"])
        x = np.linspace(-25,5,1000)
        p3 = np.polyfit(diffspepre,xvelo,1)
        y= np.polyval(p3,x)
        plt.plot(diffspepre, xvelo, "o")
        ax.plot(x, y)
        ax.set_xlabel('timeGapFolloAdj')
        ax.set_ylabel('distFolloAdj')
        plt.plot(x, y,"-")
        plt.legend(["data","polyfit"],loc="best")
        #plt.ylim(-2,2)
```

In [ ]:
```
        import scipy.integrate as integrate
        import scipy.special as special
        from sympy import *
```

In [ ]:

In [ ]:
```
        ţ= mean(vit)
        Var=varianz(vit)
        sd=math.sqrt(Var)
        print(ţ,sd,Var)
        f(x)=b0++b1x1+b2x2+....+brxr+e
```

In [ ]:
```
x=np.array(raw_data["diffSpeedPred"]).reshape((-1,1))
y=np.array(raw_data["xVelocity"])
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:
```
##linear regression
from IPython.display import display
from sympy import *
x=symbols("x")
y=symbols("y")
display( -0.47885888*x+26.74034905)
```

In [ ]:
```
#sns.lmplot(x="diffSpeedPred", y="xVelocity", hue="maneuver", data=raw_data);
```

In [ ]:

In [ ]:

In [ ]:
```
##intercept=bo,coef=b1
```

In [ ]:
```
print("bo"Model.intercept_)
```

In [ ]:
```
print("b1",Model.coef_)
```

In [ ]:
```
print(raw_data)
```

In [ ]:

In [ ]:
```
raw_data.shape
```

In [ ]:
```
for column in raw_data:
    nbr_elmt=len(raw_data[column])
    print(("Anzahl elemente in {}  :{}".format(column,nbr_elmt)))
```

In [ ]:

1 PCA Analysis

In [ ]:

In [ ]:

In [ ]:
```
from sklearn.preprocessing import MinMaxScaler
import pandas as pd
from sklearn.preprocessing import StandardScaler
#scaler=MinMaxScaler()
scaler=StandardScaler()
df=scaler.fit(raw_data)
df
```

```
In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: scaled_data

In [ ]: from sklearn.preprocessing import MinMaxScaler
```

```
import pandas as pd
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score,explained_variance_score, mean_absolute_error,mean_
from math import sqrt
import statistics as stat
from sympy import *
import scipy
from pylab import rcParams
from scipy.stats.stats import pearsonr
from sklearn.datasets import load_iris
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report
from  sklearn.decomposition import PCA as sklearnPCA
```

```
In [ ]: df1=

In [ ]: scaled_data
```

```
mean_vec=np.mean(scaled_data,axis=0)
cov_mat=(scaled_data-mean_vec).T.dot((scaled_data-mean_vec))/(scaled_data.shape[0]-1)
print("covariance matrix\n%s"%cov_mat)
```

2 Eigendecomposition en Eigenvector und eigenvalue

```
In [ ]: cov_mat=np.cov(scaled_data.T)
        eig_vals, eig_vecs=np.linalg.eig(cov_mat)
        print("Eingenvektoren \n%s" %eig_vecs)
        print("Eingenvektoren \n%s" %eig_vals)
```

3 Sortiereung der Eigenvalues

```
In [ ]: eig_pairs=[(np.abs(eig_vals[i]),eig_vecs[:,i])for i in range(len(eig_vals))]
        print("Eigenvalues in descending order:")
        for i in eig_pairs:
            print(i[0])
```

```
In [ ]: from sklearn.decomposition import PCA
        pca=PCA()
        pca.fit_transform(raw_data)
        print(pca.explained_variance_ratio_)
```

4 Explained variance

```
In [ ]: pca= PCA().fit(scaled_data)
        plt.plot(np.cumsum(pca.explained_variance_ratio_))
        plt.xlabel("Anzahl von Komponenten")
        plt.ylabel("Kumulative explained variance")
        plt.show()
```

```
In [ ]: per_var=np.round(pca.explained_variance_ratio_*100,decimals=1)
        per_var
```

```
In [ ]: pca=PCA()
        principomp = pca.fit_transform(scaled_data)
        principalDf = pd.DataFrame(data=principomp)
        principalDf
        #,columns=["PCA"+str(x)for x in range(1,len(per_var)+1)]
```

```
In [ ]: principalDf.columns
```

```
In [ ]: labels=["PCA"+str(x)for x in range(1,len(per_var)+1)]
```

```
In [ ]: plt.figure(figsize=(12,6))
        plt.bar(x=range(1,len(per_var)+1), height=per_var,tick_label=labels)
        plt.ylabel("erklärende Varinz Anteil in %")
        plt.xlabel("principal component")
        plt.title("LKL_LCR")
        plt.show()
```

```
In [ ]: wichtigkeitpca=pd.DataFrame(data=pca.explained_variance_ratio_)
        wichtigkeitpca=wichtigkeitpca.T
        print("informationsanteil in 2 PCA",wichtigkeitpca.loc[:,0:1].sum(axis=1).values)
```

```
In [ ]: fig, axes = plt.subplots(nrows=1, ncols=len(raw_data.columns), figsize=(len(raw_data.columns)*
        for index, col in enumerate(raw_data.columns):
            ax = axes[index]
            sns.boxplot(y=raw_data.loc[:,col], ax=ax)
            ax.set_ylabel(ax.get_ylabel(), fontsize='x-large')
```

```
                ax.set_xlabel("")

            #plt.tight_layout(pad=0.5)
            plt.show()
```

In [ ]:
```
def scatterPlot(xDf,yDF, algoName):
tempDF=pd.DataFrame(data=xDf.loc[:,0:1])
tempDF=pd.concat((tempDF,yDF),axis=1,join="inner")
tempDF.columns=[ "second vector PC2","First Vektor PC1","PCA2" ,'PCA3', 'PCA4', 'PCA5',
   'PCA10', 'PCA11', 'PCA12', 'PCA13', 'PCA14', 'PCA15', 'PCA16', 'PCA17','PCA18', '
sns.lmplot(x= "First Vektor PC1",y="second vector PC2",hue="Label", data=tempDF,fit_
ax=plt.gca()
ax.set_title("Separation of observation using"+algoName)
```

In [ ]:
```
scatterPlot(principalDf,scaled_data,"PCA")
```

In [ ]:
```
def scatterPlot(xDf,yDF, algoName):
tempDF=pd.DataFrame(data=xDf.loc[:,1:2])
tempDF=pd.concat((tempDF,yDF),axis=1,join="inner")
tempDF.columns=[ "second vector PC2","First Vektor PC1","PCA2" ,'PCA3', 'PCA4', 'PCA5',
    'PCA10', 'PCA11', 'PCA12', 'PCA13', 'PCA14', 'PCA15', 'PCA16', 'PCA17','PCA18', '
 sns.lmplot(x= "First Vektor PC1",y="second vector PC2",hue="Label", data=tempDF,fit_
 ax=plt.gca()
 ax.set_title("Separation of observation using"+algoName)
```

In [ ]:
```
scatterPlot(principalDf,scaled_data,"PCA")
```

In [ ]:
```
scaled_data.columns
```

In [ ]:
```
s= sns.pairplot(scaled_data[['maneuver', 'xVelocity', 'yVelocity', 'xAcceleration',
 'yAcceleration', 'class', 'distPred', 'diffSpeedPred', 'timeGapPred',
 'accPred', 'classPred', 'distFollo', 'diffSpeedFollo', 'timeGapFollo',
 'accFollo', 'classFollo', 'distPredAdj', 'diffSpeedPredAdj',
 'timeGapPredAdj', 'accPredAdj', 'classPredAdj', 'distFolloAdj',
 'diffSpeedFolloAdj', 'timeGapFolloAdj', 'accFolloAdj', 'classFolloAdj']],hue="maneuver")
```

4.1 Decision tree sur les variables Normalisees

Sans les Variables Yvelo ,Yacc,class

4.1.1 Entscheidungsbäume zur klassieren nach wichtigkeit

In [160]:
```
# split a data to x and y
x=raw_data.drop("maneuver",axis=1).values
y=raw_data["maneuver"].values
print(x.shape)
print(y.shape)
```

```
(4874, 25)
(4874,)
```

```
In [161]: dt = DecisionTreeClassifier(random_state=14, criterion="entropy",max_depth= 22)
          dt.fit(x,y)

Out[161]: DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=22,
 max_features=None, max_leaf_nodes=None,
 min_impurity_decrease=0.0, min_impurity_split=None,
 min_samples_leaf=1, min_samples_split=2,
 min_weight_fraction_leaf=0.0, presort=False, random_state=14,
 splitter='best')
```

fi_col=[] fi=[] for i,column in enumerate(raw_data.drop("maneuver",axis=1)): print("wichtigkeit der variable bezüglich der maneuver {}is: {:.3f}".format(column,dt.feature_importances_[i])) fi_col.append(column) fi.append(dt.feature_importances_[i])

```
In [ ]: fi

In [163]: fi_df=zip(fi_col,fi)
fi_df= pd.DataFrame(fi_df,columns=["variabel","Wichtigkeit"])
fi_df
#sortierung
fi_df=fi_df.sort_values("Wichtigkeit",ascending=False).reset_index()
column_tokeep=fi_df["variabel"][0:9]
fi_df
```

Out[163]:

	index	variabel	Wichtigkeit
0	1	yVelocity	0.818598
1	3	yAcceleration	0.147621
2	21	diffSpeedFolloAdj	0.013211
3	23	accFolloAdj	0.004268
4	10	distFollo	0.004066
5	2	xAcceleration	0.003910
6	6	diffSpeedPred	0.003060
7	7	timeGapPred	0.002478
8	12	timeGapFollo	0.001395
9	14	classFollo	0.001395
10	16	diffSpeedPredAdj	0.000000
11	22	timeGapFolloAdj	0.000000
12	20	distFolloAdj	0.000000
13	19	classPredAdj	0.000000
14	18	accPredAdj	0.000000
15	17	timeGapPredAdj	0.000000
16	0	xVelocity	0.000000
17	15	distPredAdj	0.000000
18	13	accFollo	0.000000
19	11	diffSpeedFollo	0.000000
20	9	classPred	0.000000
21	8	accPred	0.000000
22	5	distPred	0.000000

```
23      4              class     0.000000
24     24      classFolloAdj     0.000000
```

```
In [ ]: raw_data.columns
```

```
In [164]: new_data=raw_data[['maneuver', 'xVelocity', 'yVelocity', 'xAcceleration', 'yAcceleration',
     'class', 'distPred', 'diffSpeedPred', 'timeGapPred', 'accPred',
     'classPred', 'distFollo', 'diffSpeedFollo', 'timeGapFollo', 'accFollo',
     'classFollo', 'distPredAdj', 'diffSpeedPredAdj', 'timeGapPredAdj',
     'accPredAdj', 'classPredAdj', 'distFolloAdj', 'diffSpeedFolloAdj',
     'timeGapFolloAdj', 'accFolloAdj', 'classFolloAdj']]
```

4.1.2 Entscheidungsbäume zur klassieren nach wichtigkeit

Splitting the Raw Data_Hold_out validation https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html#sklearn.model_selection.train_test_split

```
In [165]: print(new_data.shape)
          print(new_data[column_tokeep].shape)
```

```
(4874, 26)
(4874, 9)
```

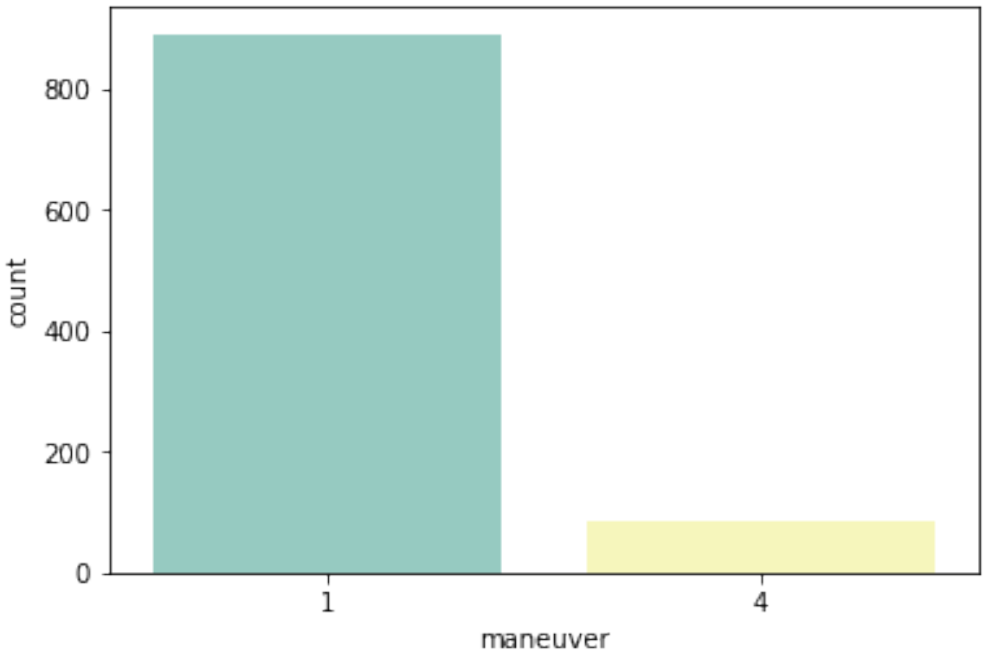
```
In [166]: x=new_data[column_tokeep].values
          x
          y=new_data["maneuver"]
          y=y.astype(int)
          y
          print(x.shape)
          print(y.shape)
```

```
(4874, 9)
(4874,)
```

```
In [184]: x_train, x_test, y_train, y_test = train_test_split(x,y, train_size=0.8,test_size=0.2,
     x_train, x_valid, y_train, y_valid = train_test_split(x_train,y_train, train_size=0.9,test_size
     print( x_train.shape)
     print( x_test.shape)
     print(x_valid.shape)
     print( y_train.shape)
     print( y_test.shape)
     print(y_valid.shape)
```

```
(3509, 9)
(975, 9)
(390, 9)
(3509,)
(975,)
(390,)
```

```
In [185]: #investigating  von alle y
          ax=sns.countplot(x=y_test,palette="Set3")
```




```
In [188]: log_reg = LogisticRegression(random_state=10,solver="lbfgs")
          log_reg.fit(x_train,y_train)

Out[188]: LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                    intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
                    penalty='l2', random_state=10, solver='lbfgs', tol=0.0001,
                    verbose=0, warm_start=False)

In [189]: y_pred=log_reg.predict(x_train)

In [190]: pred_proba=log_reg.predict_proba(x_train)
```

log_reg.coef_ x=new_data[column_tokeep] x coefi=pd.DataFrame(log_reg.coef_,columns= ["a0","a1","a2","a3","a4","a5","a6","a7","a8"]) coefi=coefi.unstack() coefi=pd.DataFrame(coefi,columns=["coef"]).values coefi

```
In [192]: i_col=pd.DataFrame(fi_col,columns=["variable"])[0:10].values
```

i_col

```
In [194]: fi_coe=zip(i_col,coefi)

          fi_coe= pd.DataFrame(fi_coe,columns=["variabel","Regressionskoefizient"])
          fi_coe
          #sortierung
```

```
#fi_coe=fi_df.sort_values("Regressionscoefizient",ascending=False).reset_index()
#column_tokeep=fi_coe["variable"][0:10]
#fi_coe
```

Out[194]:

	variabel	Regressionskoefizient
0	[xVelocity]	[-8.049239502025861]
1	[yVelocity]	[-5.635426025009274]
2	[xAcceleration]	[-0.08331992159766456]
3	[yAcceleration]	[-0.6824203020010048]
4	[class]	[0.013411781820436914]
5	[distPred]	[1.098960887634679]
6	[diffSpeedPred]	[-0.4033859262864567]
7	[timeGapPred]	[-1.0809300821345669]
8	[accPred]	[-0.3618620463474219]

4.2 Evaluation

5 Accuray on train (wie gut is the prediction)

In [182]:
```
print("the training accuracing is  :",log_reg.score(x_train,y_train))
print("the testing accuracing is  :",log_reg.score(x_test,y_test))
```

```
the training accuracing is  : 0.994300370475919
the testing accuracing is  : 0.9938461538461538
```

In [195]:
```
# classification Report
print(classification_report(y_train,y_pred))
```

```
             precision    recall  f1-score   support

          1       0.99      1.00      1.00      3229
          4       0.99      0.94      0.96       280

avg / total       0.99      0.99      0.99      3509
```

In [ ]:

In [ ]:

In [ ]:

LCL21

March 29, 2021

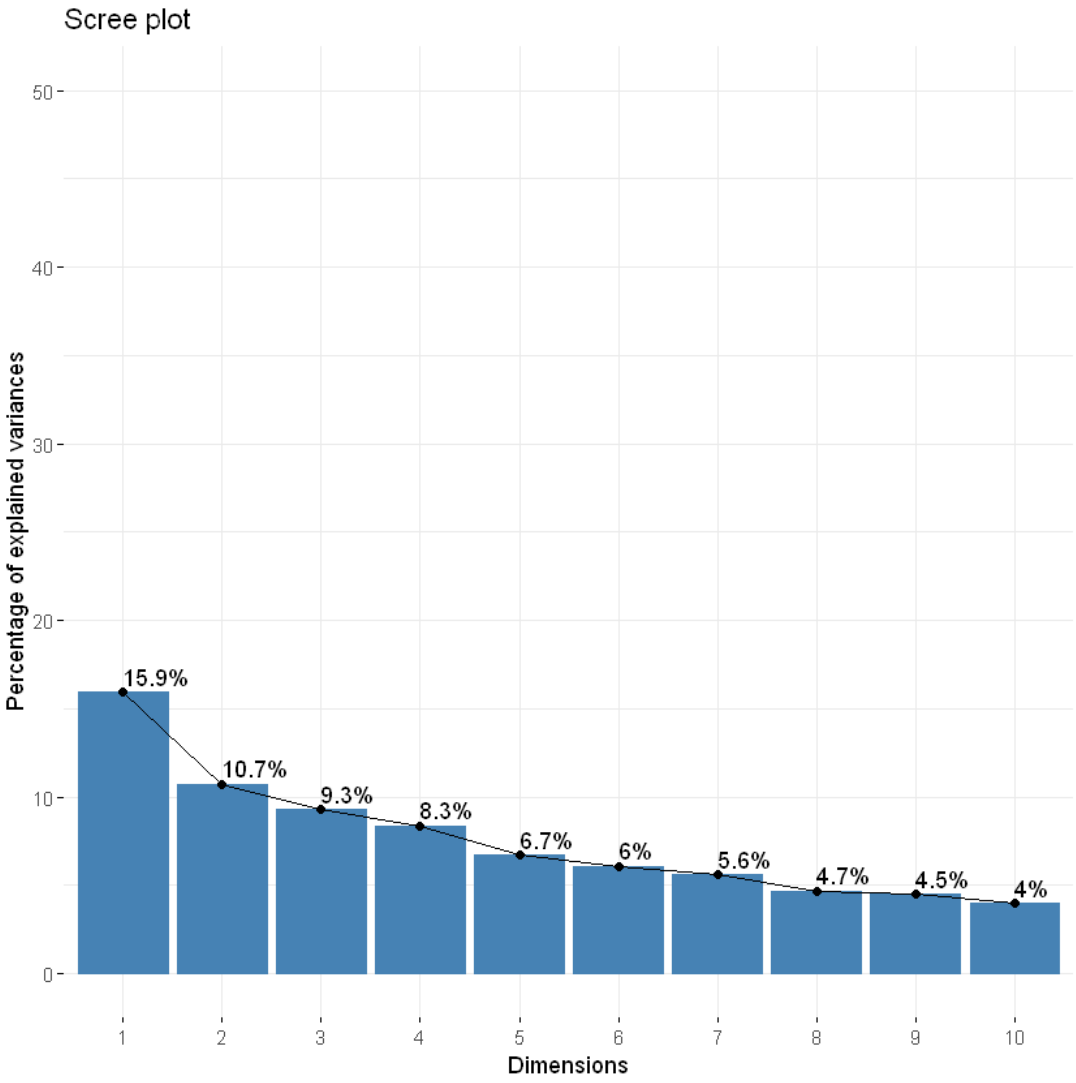
0.0.1 Principal Component analysis on HighD_dataset (R_skript)

Author: Kokou Mawuena Apedoh

```
In [6]: raw_data=read.csv("LKR_LCL1.csv")
```

raw_data=subset(raw_data, select = -c(X)) raw_data
install.packages(c("FactoMineR", "factoextra")) library("FactoMineR") library("factoextra") pca_rdata=PCA(raw_data, scale.unit = TRUE, ncp = 2, graph = TRUE) print (pca_rdata)
eigenV=get_eigenvalue(pca_rdata) eigenV

```
In [10]: fviz_eig(pca_rdata, addlabels = TRUE, ylim = c(0, 50))
```



0.1 Principal Component Analysis Results for variables

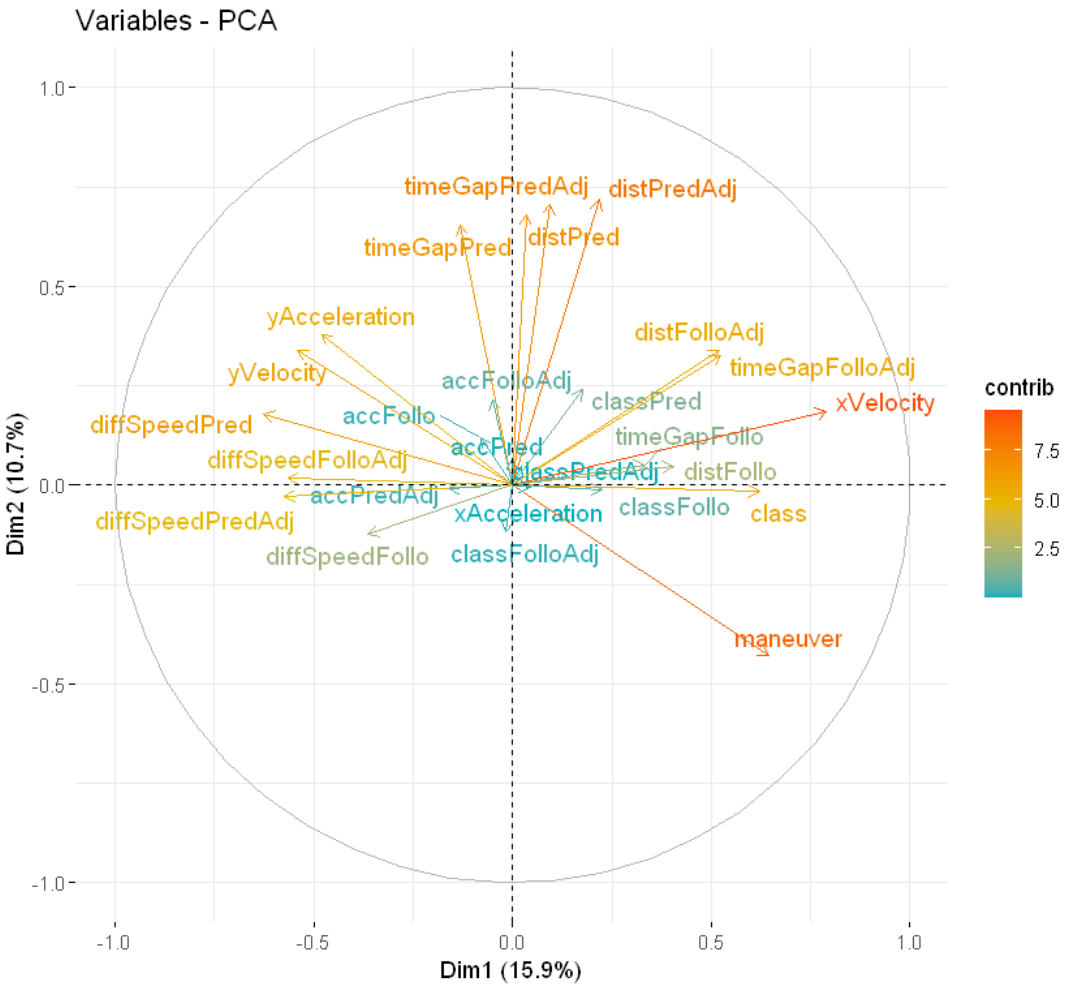
var=get_pca_var(pca_rdata) var

```
In [12]: #"Correlations between variables and dimensions"
```

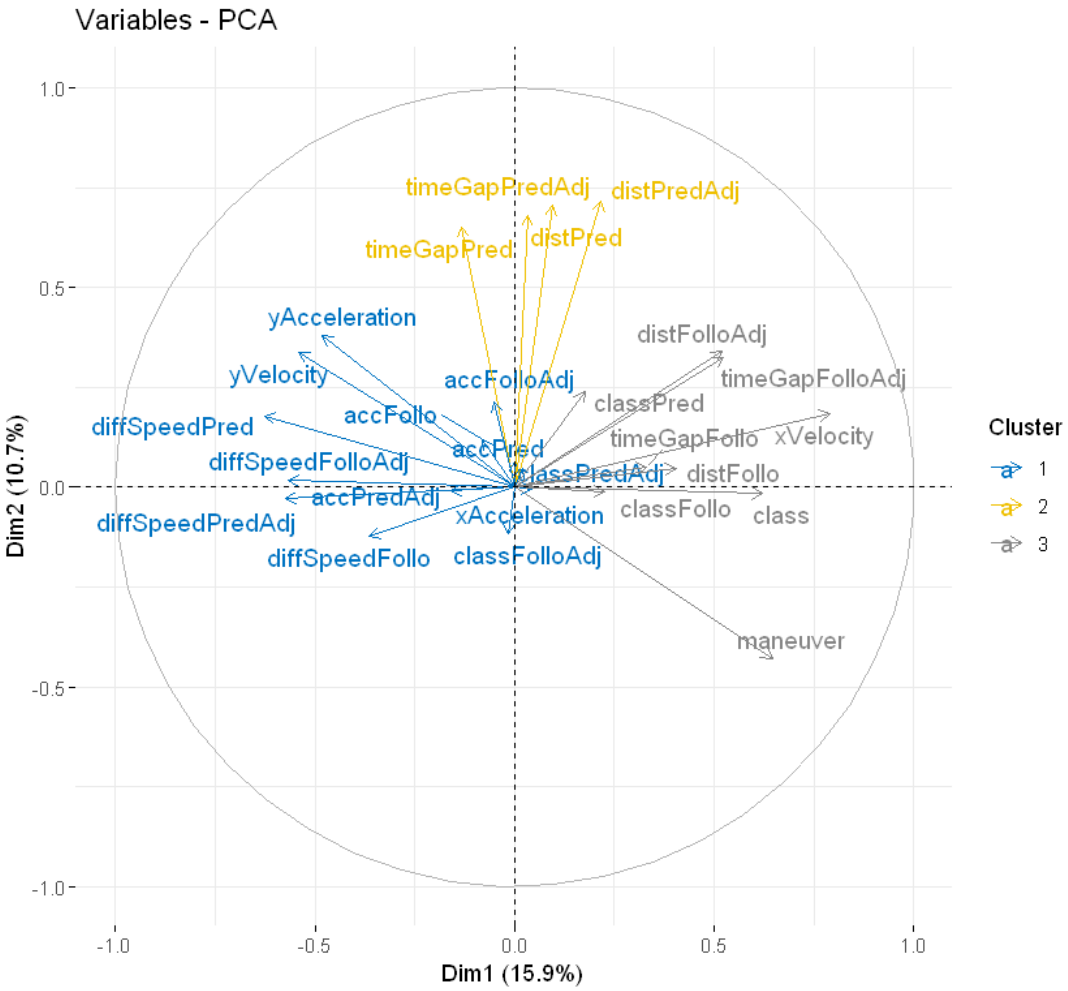
```
corre_var_PC = head (var$cor,27)
corre_var_PC
```

	Dim.1	Dim.2
maneuver	0.6462816537	-0.429540282
xVelocity	0.7904754772	0.184070103
yVelocity	-0.5402944395	0.338259176
xAcceleration	0.0447253785	-0.006910909
yAcceleration	-0.4811343489	0.379450645
class	0.6201844891	-0.015842403
distPred	0.0336229550	0.679819270
diffSpeedPred	-0.6259175248	0.178971718
timeGapPred	-0.1325977852	0.652153859
accPred	-0.0006761873	0.063487052
classPred	0.1767147036	0.239867869
distFollo	0.4045353761	0.048093259
diffSpeedFollo	-0.3633773502	-0.124471400
timeGapFollo	0.3319352738	0.050768000
accFollo	-0.0797953215	0.119805018
classFollo	0.2260799727	-0.011945219
distPredAdj	0.2177205360	0.717434788
diffSpeedPredAdj	-0.5732674762	-0.029421526
timeGapPredAdj	0.0947426184	0.705968064
accPredAdj	-0.1588305487	-0.010118429
classPredAdj	0.0214203569	0.047033091
distFolloAdj	0.5203645458	0.339305531
diffSpeedFolloAdj	-0.5660988423	0.016979427
timeGapFolloAdj	0.5243095975	0.325763523
accFolloAdj	-0.0504656392	0.213321380
classFolloAdj	-0.0171802564	-0.117882280

```
In [13]: fviz(pca_rdata, "var", repel = TRUE)
```




```
In [28]: set.seed(123)
         my.cont.var <- rnorm(26)
         # Color variables by the continuous variable
         fviz_pca_var(pca_rdata, col.var = my.cont.var,
                      gradient.cols = c("blue", "yellow", "red"),
                      legend.title = "Cont.Var")
```



0.2 Dimension description

```
In [30]: res.desc <- dimdesc(pca_rdata, axes = c(1,2), proba = 0.05)
         # Description of dimension 1
         res.desc$Dim.1
```

$quanti =

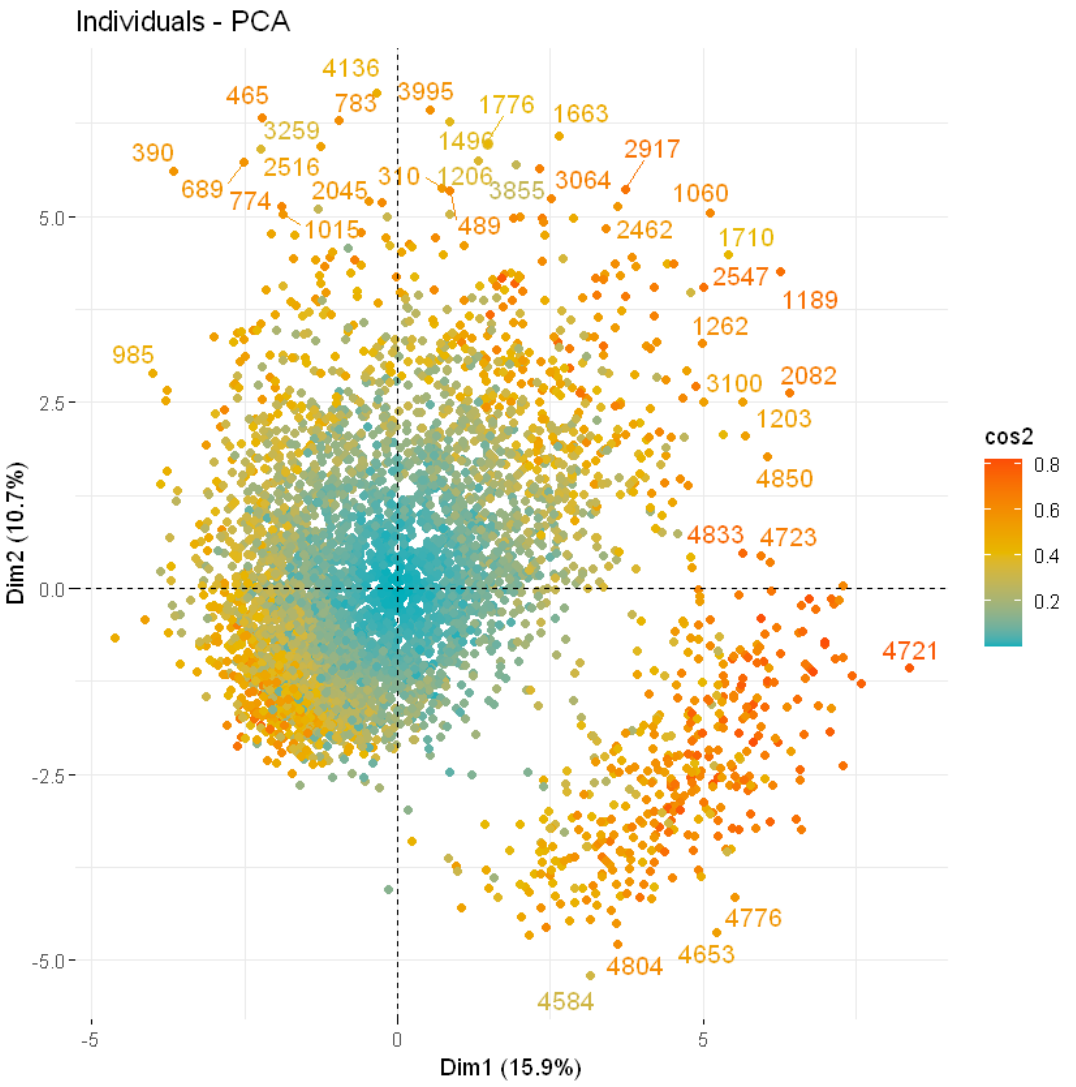
	correlation	p.value
xVelocity	0.79047548	0.000000e+00
maneuver	0.64628165	0.000000e+00
class	0.62018449	0.000000e+00
timeGapFolloAdj	0.52430960	0.000000e+00
distFolloAdj	0.52036455	0.000000e+00
distFollo	0.40453538	2.451316e-191
timeGapFollo	0.33193527	1.084284e-125
classFollo	0.22607997	1.576233e-57
distPredAdj	0.21772054	2.198263e-53
classPred	0.17671470	1.752829e-35
timeGapPredAdj	0.09474262	3.412342e-11
xAcceleration	0.04472538	1.788832e-03
distPred	0.03362295	1.890414e-02
accFolloAdj	-0.05046564	4.241993e-04
accFollo	-0.07979532	2.428789e-08
timeGapPred	-0.13259779	1.459863e-20
accPredAdj	-0.15883055	6.645109e-29
diffSpeedFollo	-0.36337735	4.771358e-152
yAcceleration	-0.48113435	6.567199e-281
yVelocity	-0.54029444	0.000000e+00
diffSpeedFolloAdj	-0.56609884	0.000000e+00
diffSpeedPredAdj	-0.57326748	0.000000e+00
diffSpeedPred	-0.62591752	0.000000e+00

res.desc$Dim.2

0.3 Plots: quality and contribution

```
In [32]: fviz_pca_ind(pca_rdata, col.ind = "cos2",
                      gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                      repel = TRUE # Avoid text overlapping (slow if many points)
                      )
```

```
Warning message:
"ggrepel: 4838 unlabeled data points (too many overlaps). Consider increasing max.overlaps"
```




```
In [33]: fviz_pca_ind(pca_rdata, pointsize = "cos2",
                      pointshape = 21, fill = "#E7B800",
                      repel = TRUE # Avoid text overlapping (slow if many points)
                      )
```

```
Warning message:
"ggrepel: 4838 unlabeled data points (too many overlaps). Consider increasing max.overlaps"
```

LCR_PCA_R

March 29, 2021

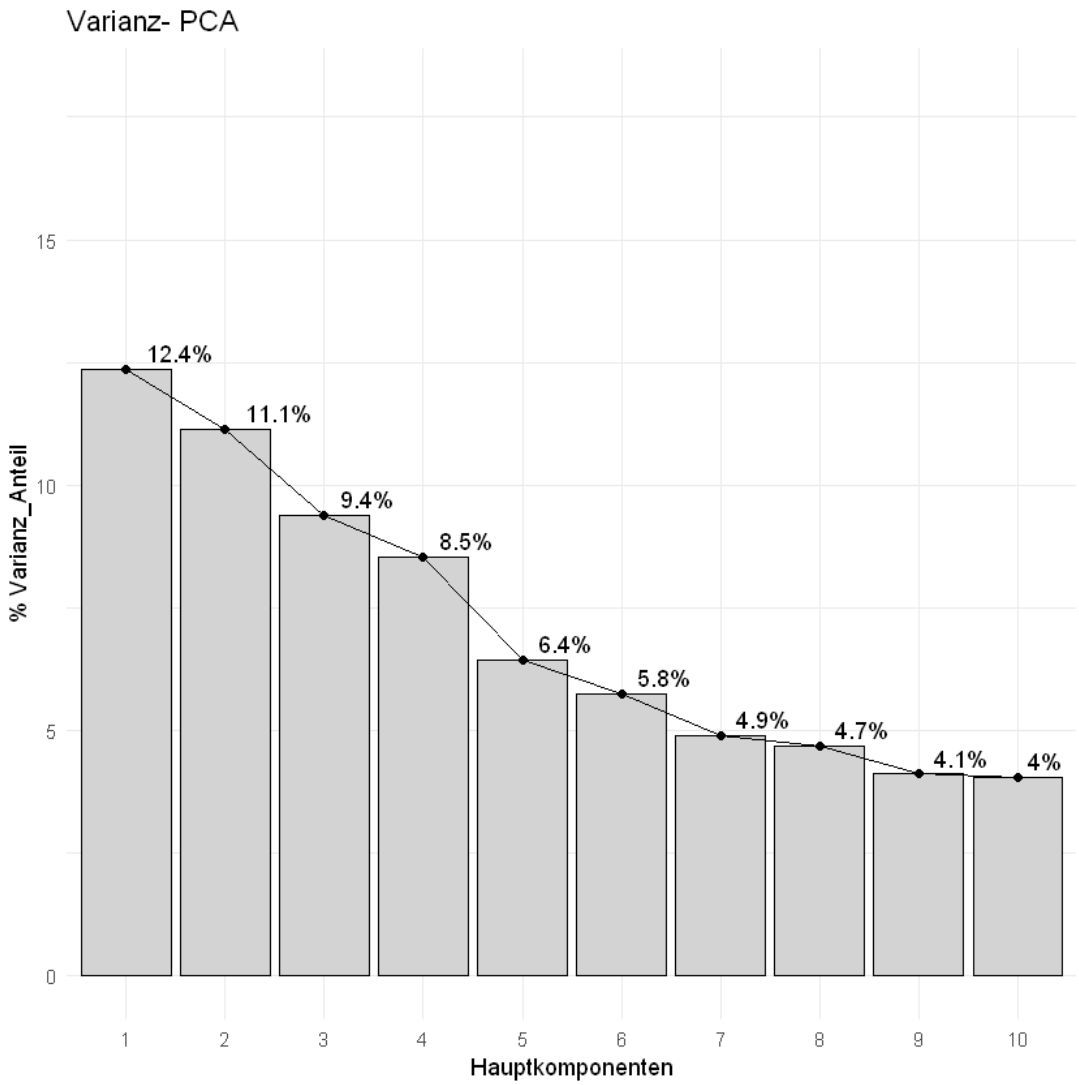
1 Principal Component analysis on HighD_dataset (R-script)

Kokou Mawuena Apedoh

```
In [16]: raw_data=read.csv("LKL_LCR.csv")
```

raw_data=subset(raw_data, select = -c(X)) raw_data
install.packages(c("FactoMineR", "factoextra")) library("FactoMineR") library("factoextra")
pca_rdata=PCA(raw_data, scale.unit = TRUE, ncp = 2, graph = TRUE) print (pca_rdata)
eigenV=get_eigenvalue(pca_rdata) eigenV

```
In [82]: p <- fviz_eig(pca_rdata, addlabels=TRUE, hjust = -0.3, barfill="lightgray",
                       barcolor ="black",
                       linecolor ="black") + ylim(0,18) +
                       theme_minimal()
         print(p)
```



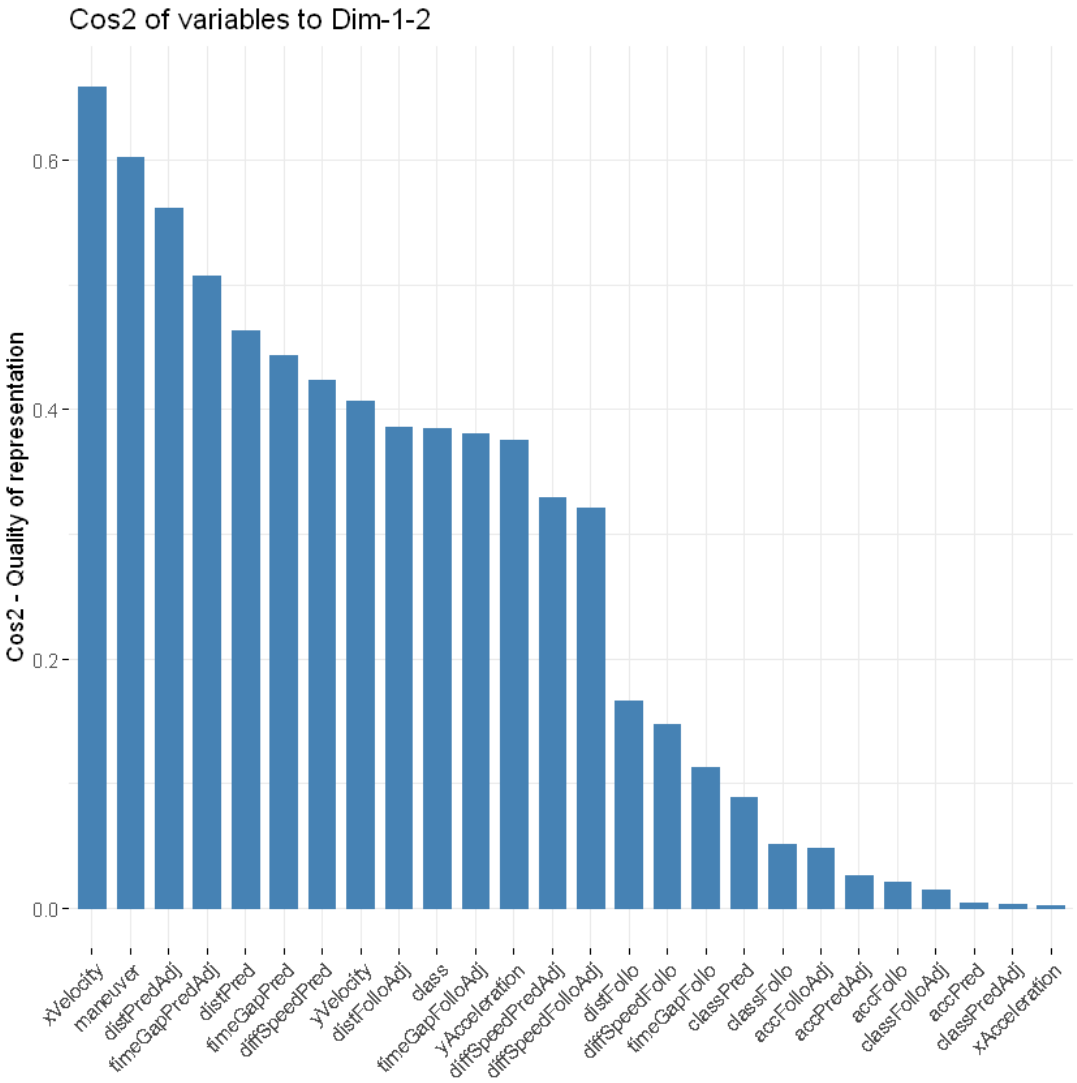
1.1 Principal Component Analysis Results for variables

var=get_pca_var(pca_rdata) var

```
In [87]: #"Correlations between variables and dimensions"
         corre_var_PC = head (var$cor,27)
         corre_var_PC
```

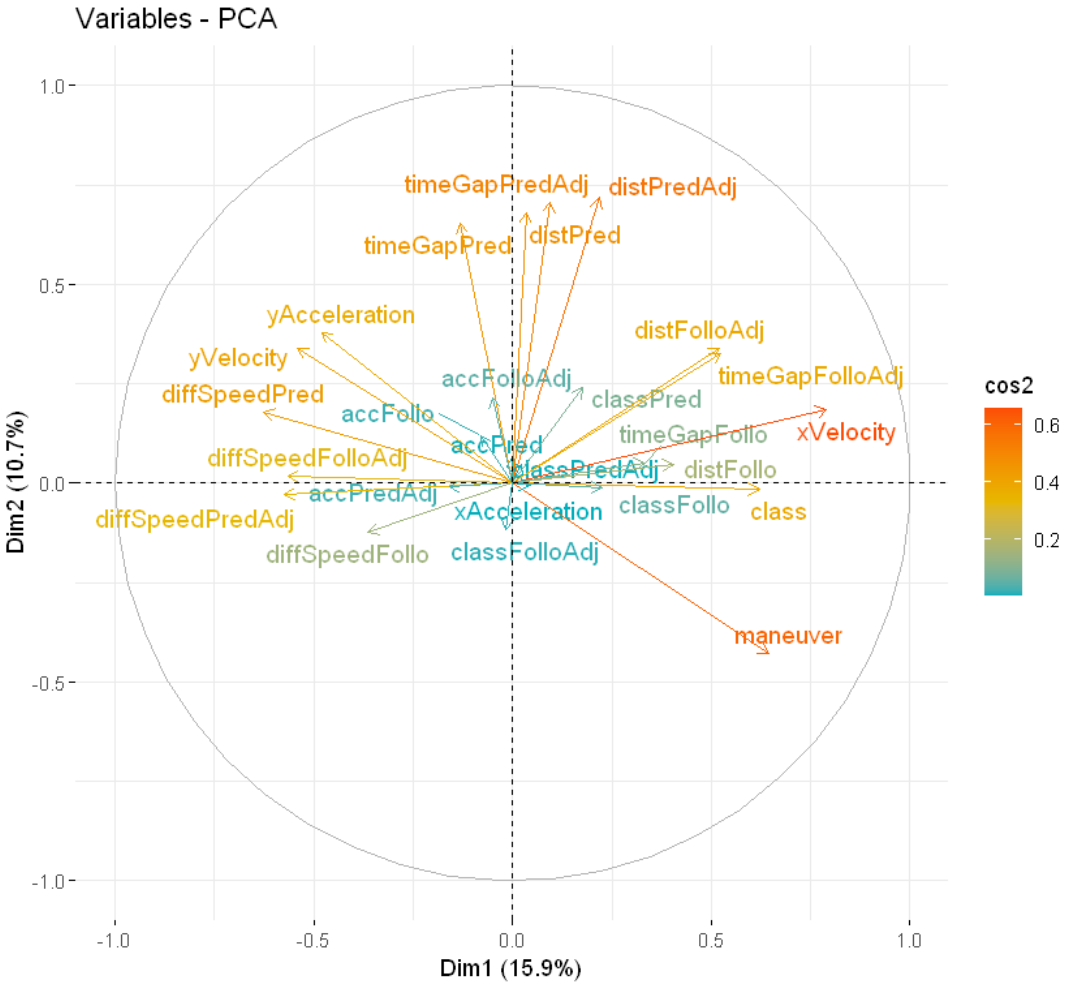
	Dim.1	Dim.2
maneuver	0.09646091	0.599747813
xVelocity	0.81039434	-0.007950761
yVelocity	0.02092117	-0.155948166
xAcceleration	-0.25222447	0.401610790
yAcceleration	0.05913447	0.141062821
class	0.09277139	-0.216848471
distPred	0.53025705	0.559428931
diffSpeedPred	-0.40611500	0.383626833
timeGapPred	0.42729462	0.598403189
accPred	-0.31811588	0.256977770
classPred	0.09433757	0.067744045
distFollo	0.62650525	-0.150296702
diffSpeedFollo	-0.25022864	-0.109428900
timeGapFollo	0.57655410	-0.134649932
accFollo	-0.02733246	0.411515996
classFollo	-0.02918851	-0.001866533
distPredAdj	0.25221219	0.536576166
diffSpeedPredAdj	-0.58161656	0.346074156
timeGapPredAdj	0.12647225	0.547041513
accPredAdj	-0.24759030	0.297551293
classPredAdj	0.09241447	0.239859403
distFolloAdj	0.16319618	-0.363489446
diffSpeedFolloAdj	-0.61990460	-0.126588662
timeGapFolloAdj	0.14731972	-0.337818560
accFolloAdj	-0.14653232	0.193951935
classFolloAdj	-0.04456413	-0.149333079

```
In [96]: fviz(pca_rdata, "var",
                      gradient.cols = c("white", "#2E9FDF", "#FC4E07" ),
                      repel = TRUE )
```



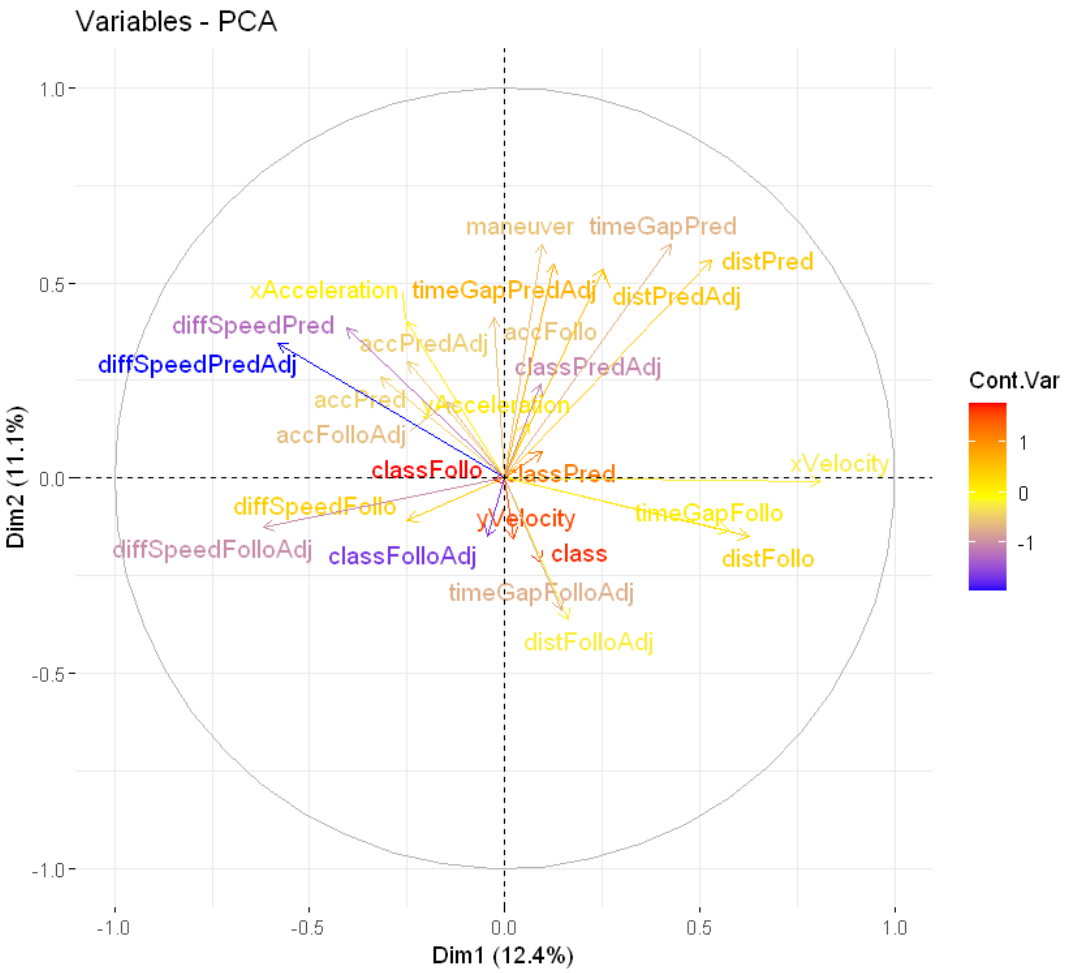
library("corrplot") corrplot(var$cos2, is.corr=FALSE)

```
In [17]: fviz_pca_var(pca_rdata, col.var = "cos2",
                      gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                      repel = TRUE # Avoid text overlapping
                      )
```

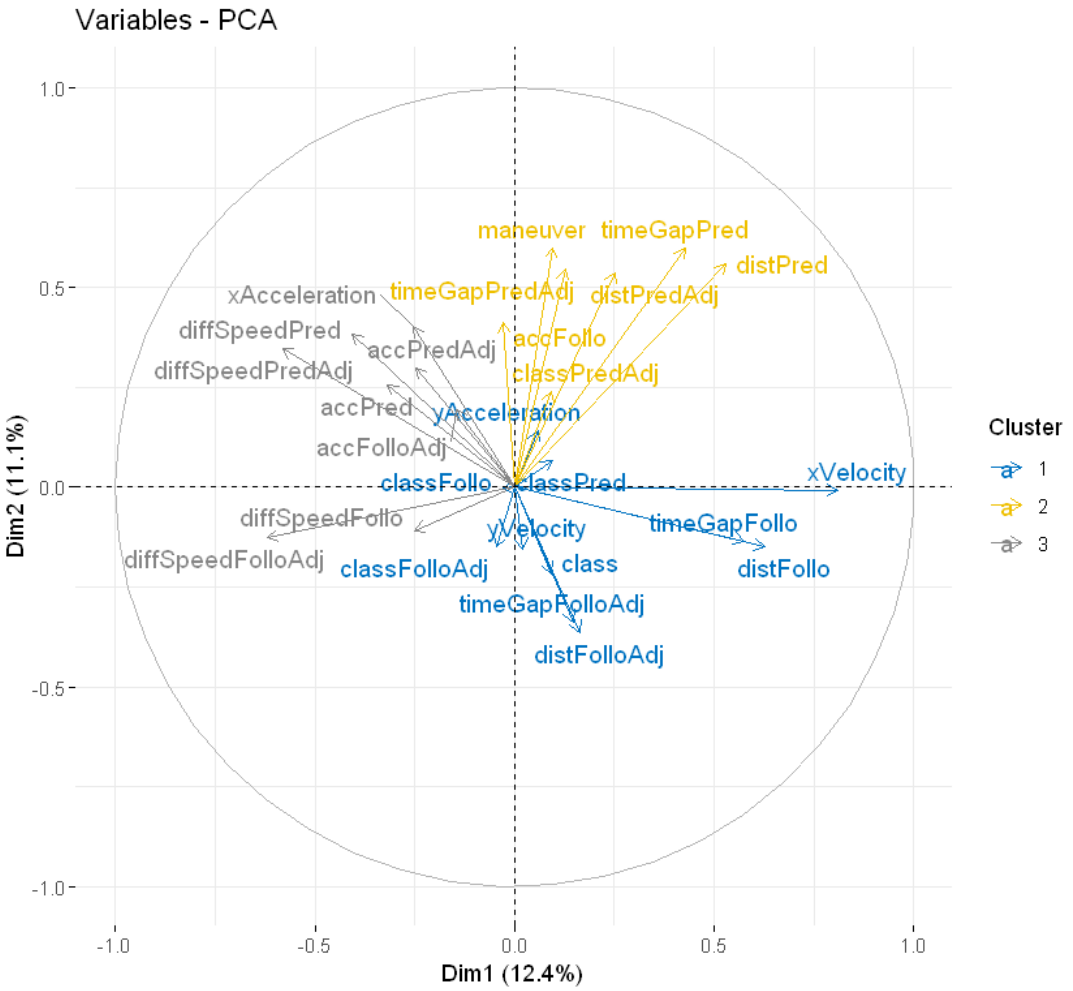


head(var$contrib, 27)
corrplot(var$contrib, is.corr=FALSE)

```
In [22]: # Contributions of variables to PC1
         fviz_contrib(pca_rdata, choice = "var", axes = 1, top = 10)
         # Contributions of variables to PC2
         fviz_contrib(pca_rdata, choice = "var", axes = 2, top = 10)
```




```
In [98]: set.seed(123)
         res.km <- kmeans(var$coord, centers = 3, nstart = 25)
         grp <- as.factor(res.km$cluster)
         # Color variables by groups
         fviz_pca_var(pca_rdata, col.var = grp,
                      palette = c("#0073C2FF", "#EFC000FF", "#868686FF"),
                      legend.title = "Cluster", repel = TRUE)
```



1.2 Dimension description

```
In [30]: res.desc <- dimdesc(pca_rdata, axes = c(1,2), proba = 0.05)
         # Description of dimension 1
         res.desc$Dim.1
```

$quanti =

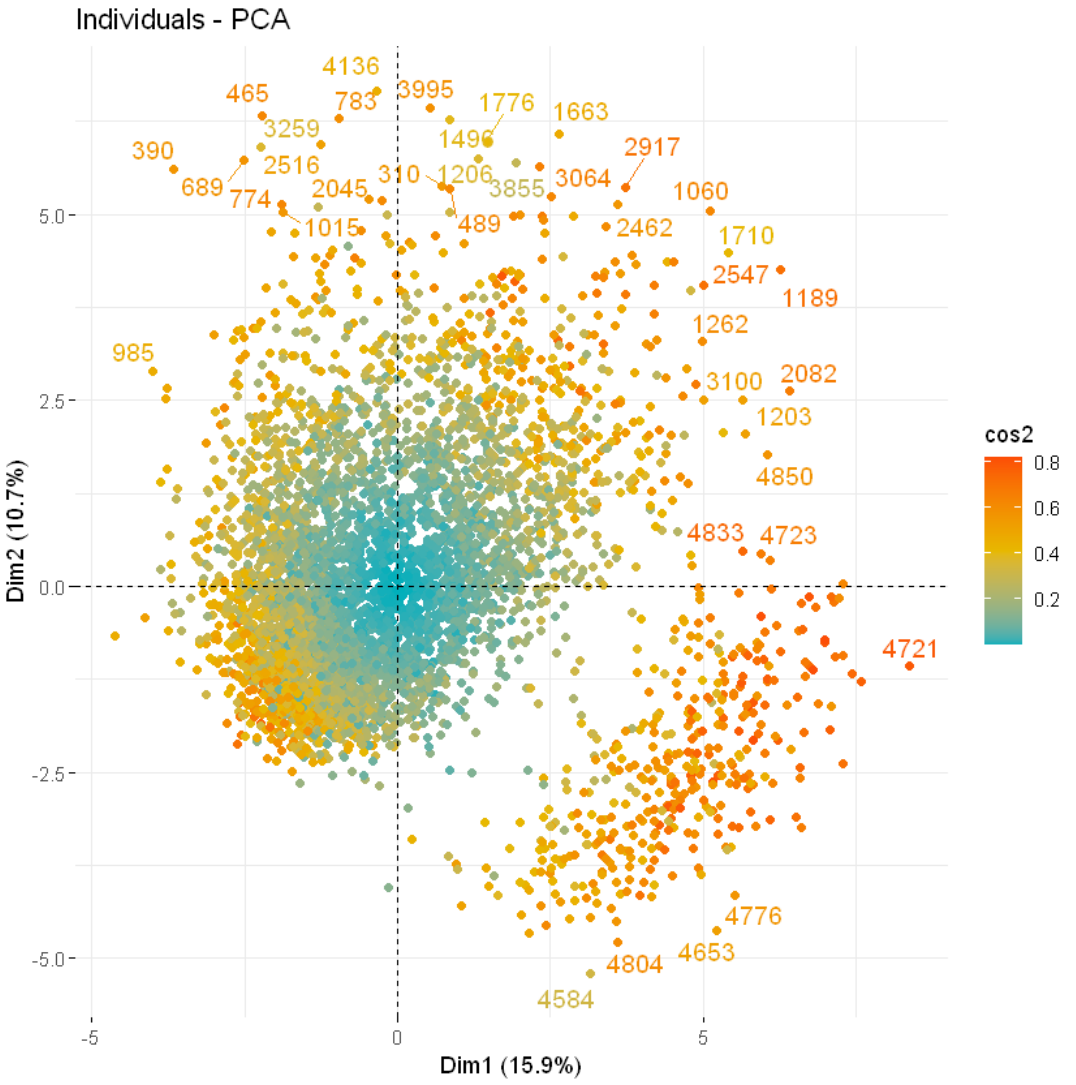
	correlation	p.value
xVelocity	0.79047548	0.000000e+00
maneuver	0.64628165	0.000000e+00
class	0.62018449	0.000000e+00
timeGapFolloAdj	0.52430960	0.000000e+00
distFolloAdj	0.52036455	0.000000e+00
distFollo	0.40453538	2.451316e-191
timeGapFollo	0.33193527	1.084284e-125
classFollo	0.22607997	1.576233e-57
distPredAdj	0.21772054	2.198263e-53
classPred	0.17671470	1.752829e-35
timeGapPredAdj	0.09474262	3.412342e-11
xAcceleration	0.04472538	1.788832e-03
distPred	0.03362295	1.890414e-02
accFolloAdj	-0.05046564	4.241993e-04
accFollo	-0.07979532	2.428789e-08
timeGapPred	-0.13259779	1.459863e-20
accPredAdj	-0.15883055	6.645109e-29
diffSpeedFollo	-0.36337735	4.771358e-152
yAcceleration	-0.48113435	6.567199e-281
yVelocity	-0.54029444	0.000000e+00
diffSpeedFolloAdj	-0.56609884	0.000000e+00
diffSpeedPredAdj	-0.57326748	0.000000e+00
diffSpeedPred	-0.62591752	0.000000e+00

res.desc$Dim.2

1.3 Plots: quality and contribution

```
In [32]: fviz_pca_ind(pca_rdata, col.ind = "cos2",
                      gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                      repel = TRUE # Avoid text overlapping (slow if many points)
                      )
```

```
Warning message:
"ggrepel: 4838 unlabeled data points (too many overlaps). Consider increasing max.overlaps"
```




```
In [33]: fviz_pca_ind(pca_rdata, pointsize = "cos2",
                      pointshape = 21, fill = "#E7B800",
                      repel = TRUE # Avoid text overlapping (slow if many points)
                      )
```

```
Warning message:
"ggrepel: 4838 unlabeled data points (too many overlaps). Consider increasing max.overlaps"
```

LCR21

March 29, 2021

0.0.1 Principal Component analysis on HighD_dataset (R_Skript)

Author: Kokou Mawuena Apedoh

In [77]:
```
raw_data=read.csv("LKL_LCR.csv")
```

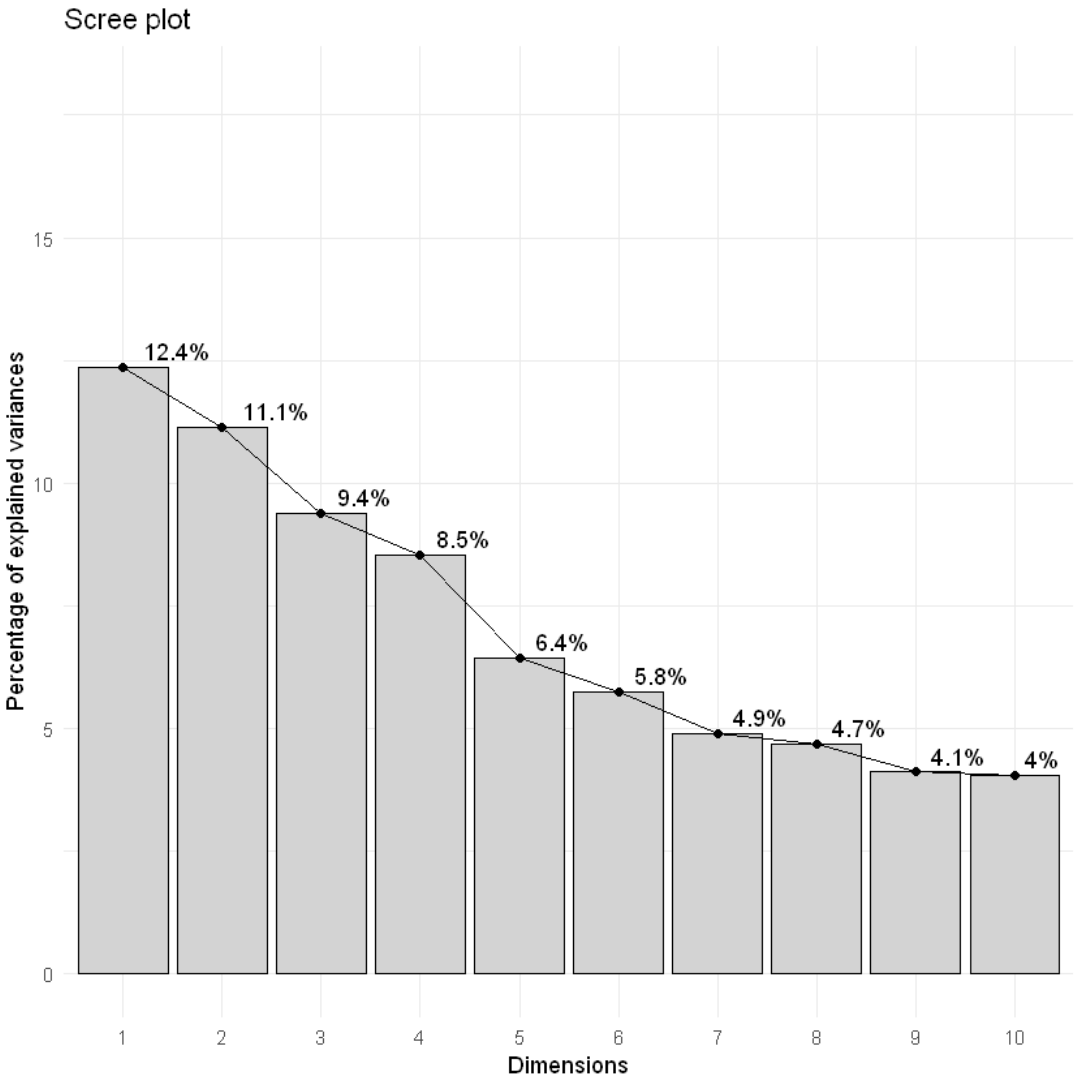
raw_data=subset(raw_data, select = -c(X)) raw_data
install.packages(c("FactoMineR", "factoextra")) library("FactoMineR") library("factoextra") pca_rdata=PCA(raw_data, scale.unit = TRUE, ncp = 2, graph = TRUE) print (pca_rdata)

In [ ]:
```
```

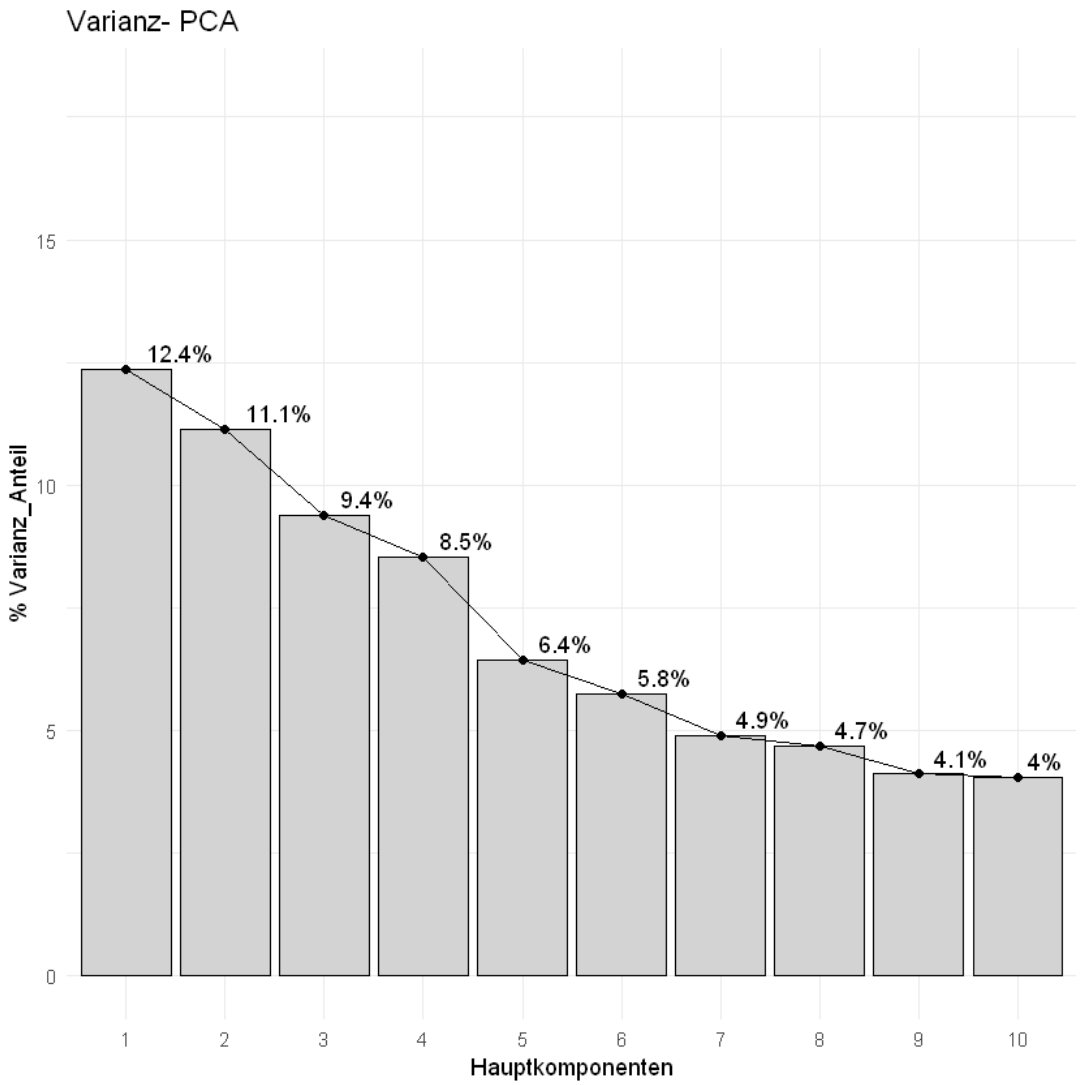
In [85]:
```
eigenV=get_eigenvalue(pca_rdata)
eigenV
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	3.213919665	12.36122948	12.36123
Dim.2	2.898373852	11.14759174	23.50882
Dim.3	2.440941052	9.38823481	32.89706
Dim.4	2.218496611	8.53267927	41.42974
Dim.5	1.674105503	6.43886732	47.86860
Dim.6	1.495471795	5.75181459	53.62042
Dim.7	1.274738100	4.90283884	58.52326
Dim.8	1.218147836	4.68518398	63.20844
Dim.9	1.067891172	4.10727374	67.31571
Dim.10	1.051064200	4.04255462	71.35827
Dim.11	0.964827744	3.71087594	75.06914
Dim.12	0.910726371	3.50279373	78.57194
Dim.13	0.834295746	3.20882979	81.78077
Dim.14	0.803091558	3.08881368	84.86958
Dim.15	0.701220207	2.69700079	87.56658
Dim.16	0.673525137	2.59048130	90.15706
Dim.17	0.625668569	2.40641757	92.56348
Dim.18	0.601969406	2.31526695	94.87875
Dim.19	0.543904215	2.09193929	96.97069
Dim.20	0.378546707	1.45594887	98.42664
Dim.21	0.225873020	0.86874239	99.29538
Dim.22	0.148160493	0.56984805	99.86523
Dim.23	0.010844769	0.04171065	99.90694
Dim.24	0.008972002	0.03450770	99.94145
Dim.25	0.008336069	0.03206181	99.97351
Dim.26	0.006888201	0.02649308	100.00000

```
In [82]: p <- fviz_eig(pca_rdata, addlabels=TRUE, hjust = -0.3, barfill="lightgray",
                       barcolor ="black",
                       linecolor ="black") + ylim(0,18) +
                       theme_minimal()
         print(p)
```




```
In [83]: p + labs(title = "Varianz- PCA",
                  x = "Hauptkomponenten", y = "% Varianz_Anteil")
```



0.1 Principal Component Analysis Results for variables

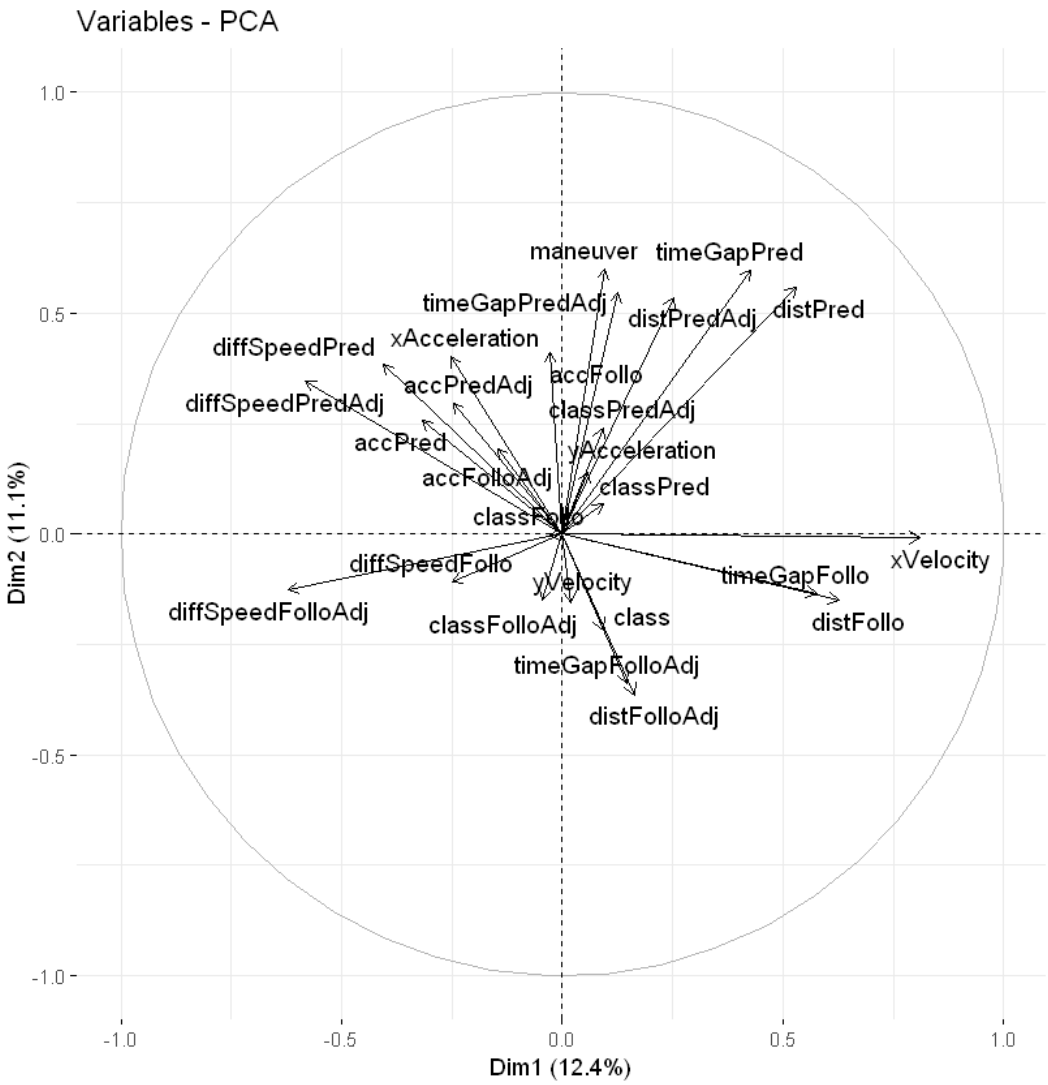
```
In [86]: var=get_pca_var(pca_rdata)
         var

Principal Component Analysis Results for variables
 ===================================================
  Name       Description
1 "$coord"   "Coordinates for the variables"
2 "$cor"     "Correlations between variables and dimensions"
3 "$cos2"    "Cos2 for the variables"
4 "$contrib" "contributions of the variables"
```

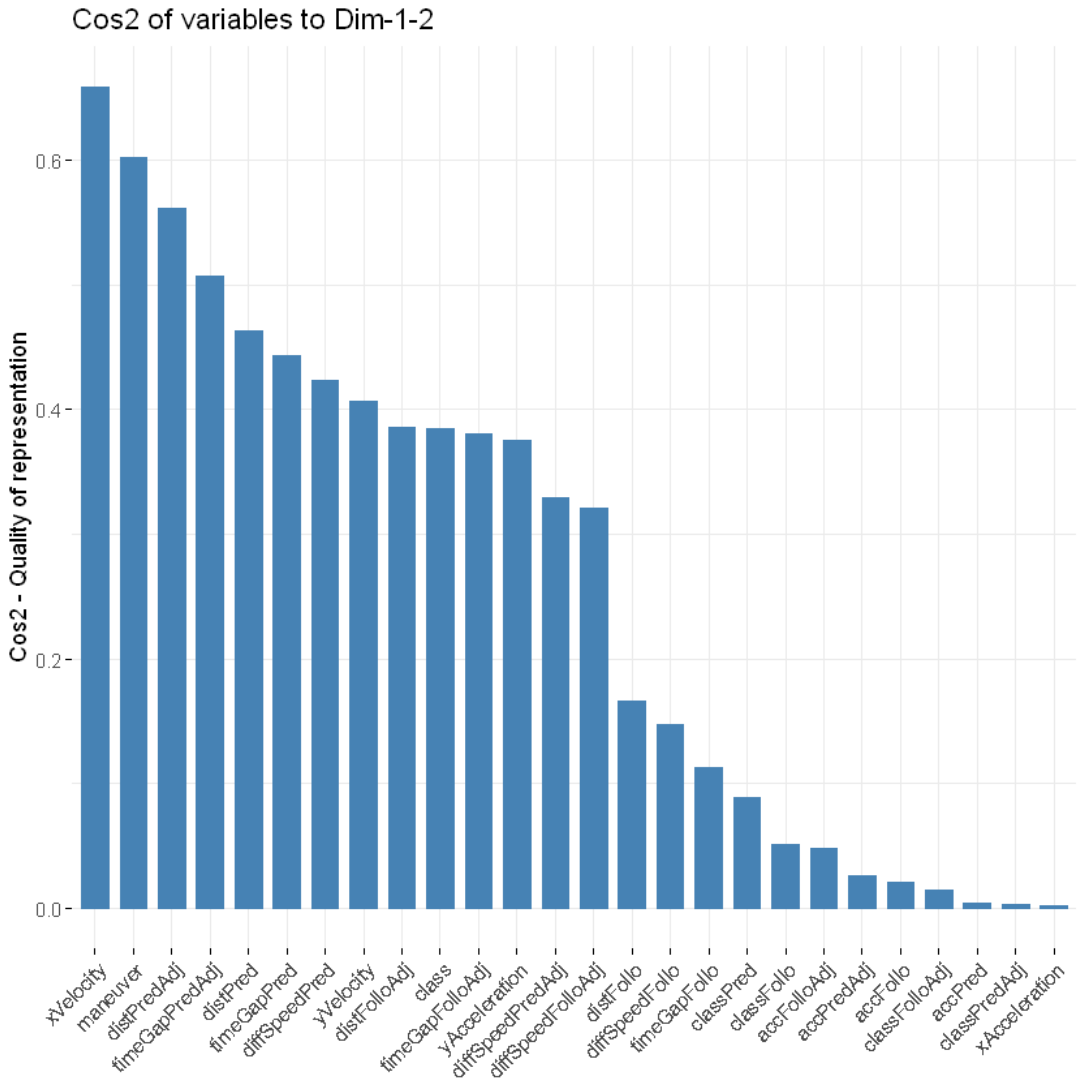
```
In [87]: #"Correlations between variables and dimensions"
         corre_var_PC = head (var$cor,27)
         corre_var_PC
```

	Dim.1	Dim.2
maneuver	0.09646091	0.599747813
xVelocity	0.81039434	-0.007950761
yVelocity	0.02092117	-0.155948166
xAcceleration	-0.25222447	0.401610790
yAcceleration	0.05913447	0.141062821
class	0.09277139	-0.216848471
distPred	0.53025705	0.559428931
diffSpeedPred	-0.40611500	0.383626833
timeGapPred	0.42729462	0.598403189
accPred	-0.31811588	0.256977770
classPred	0.09433757	0.067744045
distFollo	0.62650525	-0.150296702
diffSpeedFollo	-0.25022864	-0.109428900
timeGapFollo	0.57655410	-0.134649932
accFollo	-0.02733246	0.411515996
classFollo	-0.02918851	-0.001866533
distPredAdj	0.25221219	0.536576166
diffSpeedPredAdj	-0.58161656	0.346074156
timeGapPredAdj	0.12647225	0.547041513
accPredAdj	-0.24759030	0.297551293
classPredAdj	0.09241447	0.239859403
distFolloAdj	0.16319618	-0.363489446
diffSpeedFolloAdj	-0.61990460	-0.126588662
timeGapFolloAdj	0.14731972	-0.337818560
accFolloAdj	-0.14653232	0.193951935
classFolloAdj	-0.04456413	-0.149333079

```
In [96]: fviz(pca_rdata, "var",
                      gradient.cols = c("white", "#2E9FDF", "#FC4E07" ),
                      repel = TRUE )
```

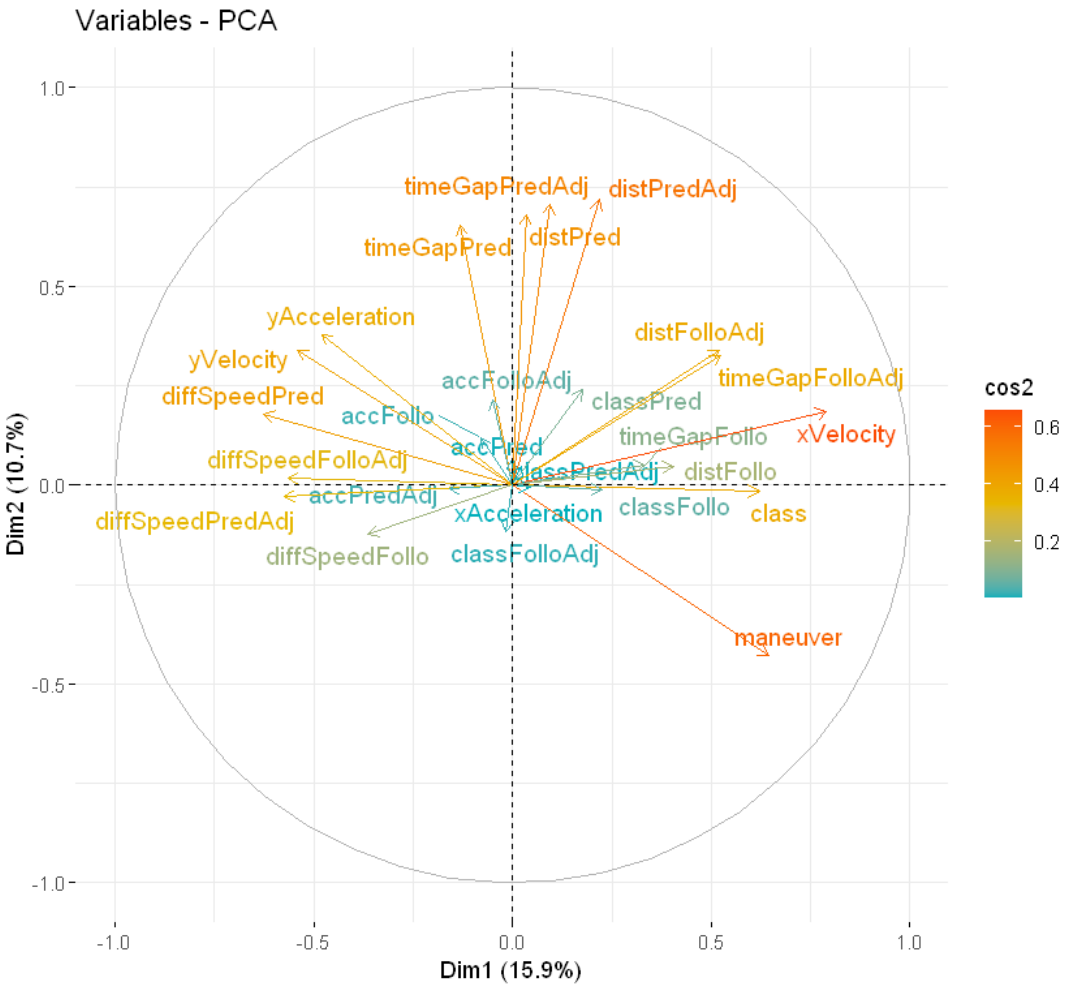



```
In [15]: fviz_cos2(pca_rdata, choice = "var", axes = 1:2)
```




```
In [16]: library("corrplot")
         corrplot(var$cos2, is.corr=FALSE)
```

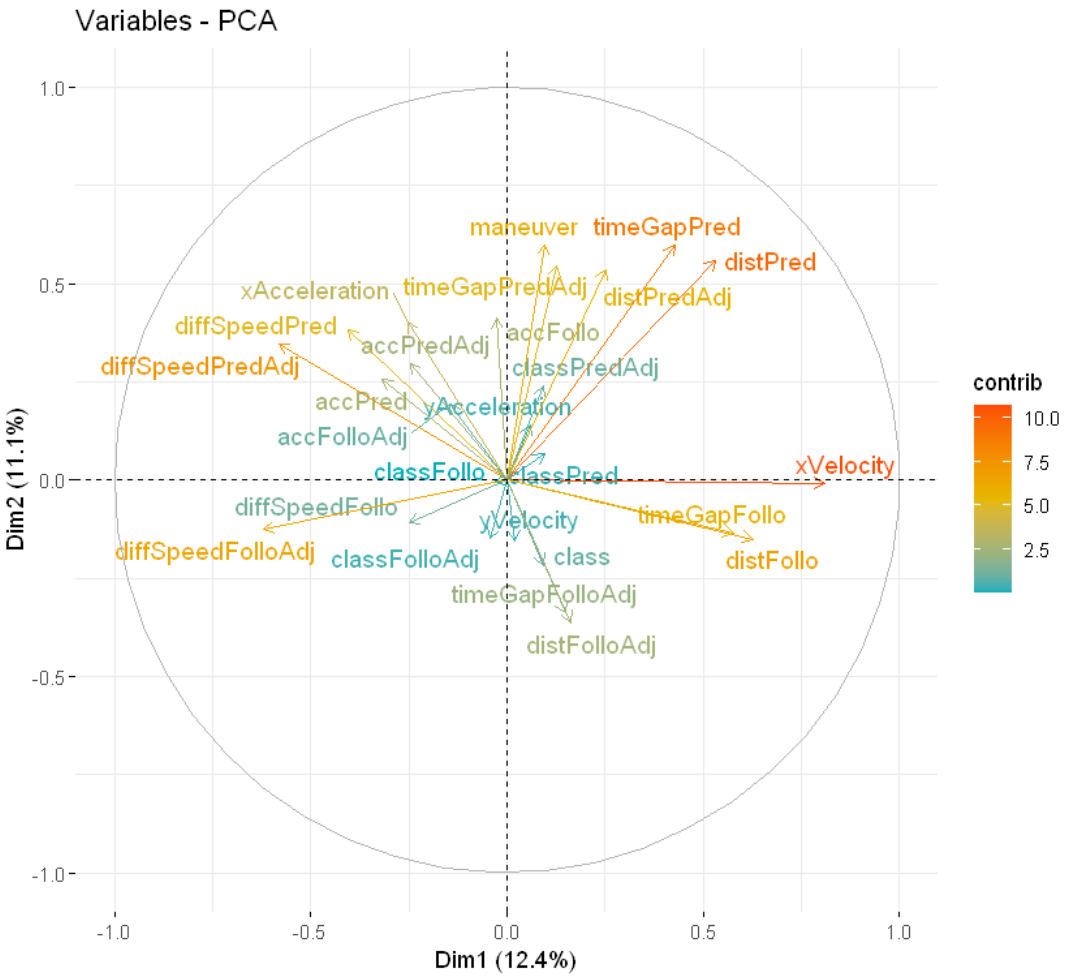
```
Warning message:
"package 'corrplot' was built under R version 3.6.3"corrplot 0.84 loaded
```



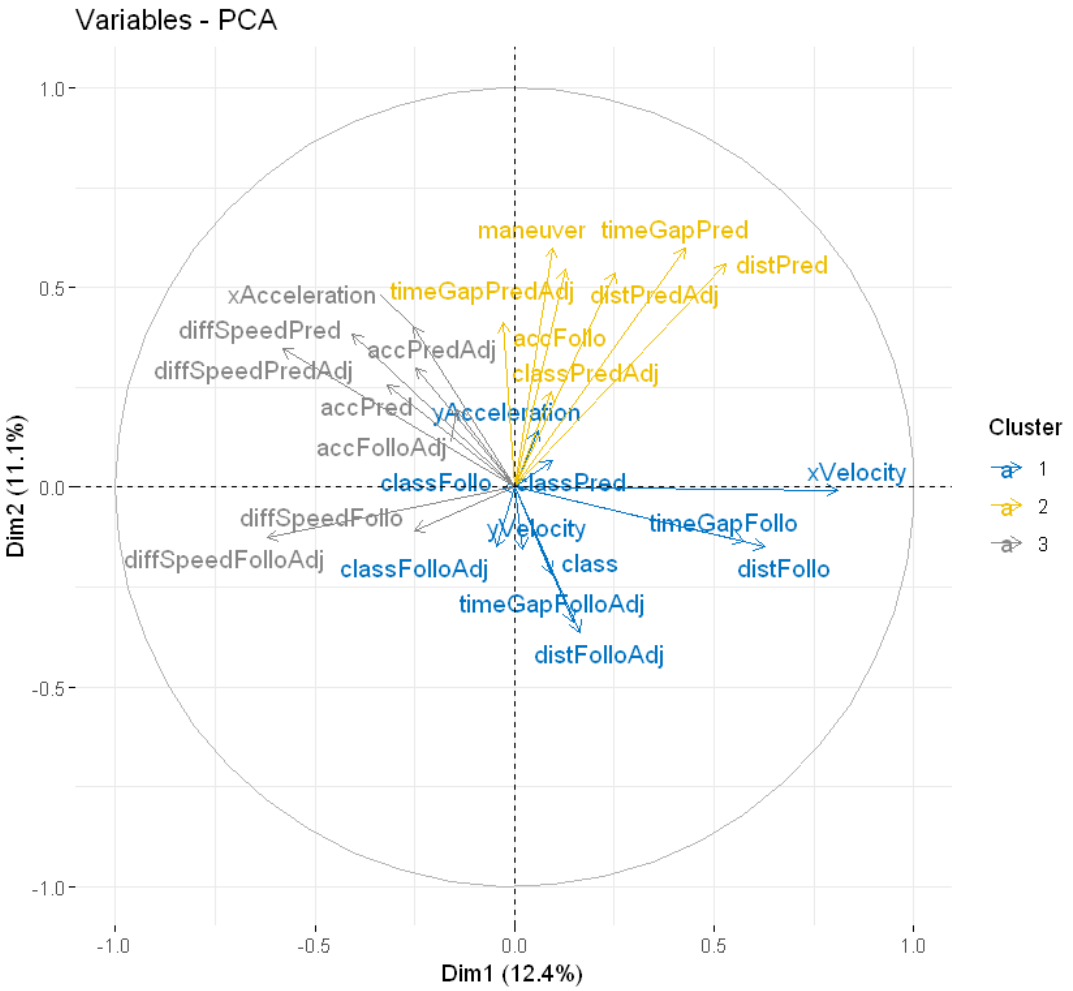
In [19]: `head(var$contrib, 27)`

	Dim.1	Dim.2
maneuver	1.009541e+01	6.612488919
xVelocity	1.510279e+01	1.214293513
yVelocity	7.055720e+00	4.100684308
xAcceleration	4.834910e-02	0.001711700
yAcceleration	5.595167e+00	5.160214463
class	9.296558e+00	0.008994960
distPred	2.732449e-02	16.563194586
diffSpeedPred	9.469228e+00	1.147957941
timeGapPred	4.249647e-01	15.242538449
accPred	1.105133e-05	0.144453307
classPred	7.547893e-01	2.062060081
distFollo	3.955427e+00	0.082894472
diffSpeedFollo	3.191509e+00	0.555259902
timeGapFollo	2.663098e+00	0.092371347
accFollo	1.538988e-01	0.514407369
classFollo	1.235391e+00	0.005113819
distPredAdj	1.145722e+00	18.446841892
diffSpeedPredAdj	7.943190e+00	0.031023269
timeGapPredAdj	2.169557e-01	17.861884332
accPredAdj	6.097453e-01	0.003669301
classPredAdj	1.109006e-02	0.079280102
distFolloAdj	6.544791e+00	4.126093267
diffSpeedFolloAdj	7.745775e+00	0.010332448
timeGapFolloAdj	6.644403e+00	3.803312996
accFolloAdj	6.155622e-02	1.630894737
classFolloAdj	7.134107e-03	0.498028520

```
In [20]: corrplot(var$contrib, is.corr=FALSE)
```




```
In [97]: set.seed(123)
         my.cont.var <- rnorm(26)
         # Color variables by the continuous variable
         fviz_pca_var(pca_rdata, col.var = my.cont.var,
                      gradient.cols = c("blue", "yellow", "red"),
                      legend.title = "Cont.Var", repel = TRUE)
```



0.2 Dimension description

```
In [30]: res.desc <- dimdesc(pca_rdata, axes = c(1,2), proba = 0.05)
         # Description of dimension 1
         res.desc$Dim.1
```

$quanti =

	correlation	p.value
xVelocity	0.79047548	0.000000e+00
maneuver	0.64628165	0.000000e+00
class	0.62018449	0.000000e+00
timeGapFolloAdj	0.52430960	0.000000e+00
distFolloAdj	0.52036455	0.000000e+00
distFollo	0.40453538	2.451316e-191
timeGapFollo	0.33193527	1.084284e-125
classFollo	0.22607997	1.576233e-57
distPredAdj	0.21772054	2.198263e-53
classPred	0.17671470	1.752829e-35
timeGapPredAdj	0.09474262	3.412342e-11
xAcceleration	0.04472538	1.788832e-03
distPred	0.03362295	1.890414e-02
accFolloAdj	-0.05046564	4.241993e-04
accFollo	-0.07979532	2.428789e-08
timeGapPred	-0.13259779	1.459863e-20
accPredAdj	-0.15883055	6.645109e-29
diffSpeedFollo	-0.36337735	4.771358e-152
yAcceleration	-0.48113435	6.567199e-281
yVelocity	-0.54029444	0.000000e+00
diffSpeedFolloAdj	-0.56609884	0.000000e+00
diffSpeedPredAdj	-0.57326748	0.000000e+00
diffSpeedPred	-0.62591752	0.000000e+00

res.desc$Dim.2

0.3 Plots: quality and contribution

```
In [32]: fviz_pca_ind(pca_rdata, col.ind = "cos2",
                      gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                      repel = TRUE # Avoid text overlapping (slow if many points)
                      )
```

```
Warning message:
"ggrepel: 4838 unlabeled data points (too many overlaps). Consider increasing max.overlaps"
```