

Hausarbeit

Einflussfaktoren auf die Schulleistungen im Rahmen der Oberschule

MADS2100 Reporting und Visualisierung 23oB

Leon Henne

Köln, den 22. März 2024

Betreut durch Dr. Robert Stahlbock

Inhaltsverzeichnis

Tabellenverzeichnis	I
Abbildungsverzeichnis	II
1 Problemstellung	1
2 Zielsetzung	2
2.1 Untersuchter Datensatz	2
2.2 Forschungsfragen	3
3 Grundlagen	4
3.1 Data Mining im Bildungsbereich	4
3.2 Business Analytics	4
4 Untersuchung der Forschungsfrage	5
4.1 Individuelle und strukturelle Faktoren	5
4.2 Soziales Umfeld	8
4.3 Individuelle Leistungsbereitschaft	10
5 Fazit und Ausblick	13
Literatur	14
6 Anhang	15

Tabellenverzeichnis

2.1	Kurzbeschreibung der im Datensatz enthaltenen Merkmale	3
6.1	Detaillierte Darstellung des untersuchten Datensatzes	16

Abbildungsverzeichnis

4.1	Einfluss von Alkoholkonsum auf die Lernleistung	6
4.2	Zusammenhang struktureller Faktoren und Lernleistungen	7
4.3	Einfluss der Familiengröße und dessen Beziehungsqualität auf die Lernleistung	8
4.4	Lernleistungseinfluss des Partnerstatus und sozialer Aktivitäten	9
4.5	Einfluss von Lernunterstützungen zu den Lernleistungen von Schülern	11
4.6	Lernleistung anhand der Lernzeit und Anzahl der Schulfehltag	12

1 Problemstellung

Der für jede Nation erstrebenswerte langfristige ökonomische Fortschritt wird unter anderem stark durch das vorherrschende Bildungsniveau beeinflusst (Cortez und Silva, 2008, S. 1). Um dieses sich ergebende Bildungsniveau stärker zu durchleuchten und final zu verbessern, wird zur Unterstützung der Schüler und Lehrkräfte die Modellierung von Schulleistungen eingesetzt (Cortez und Silva, 2008, S. 1). So kann eine zeitabhängige Vorhersage der Leistungen, lernschwächere Schüler detektieren und damit Lehrkräfte frühzeitig befähigen, mit entsprechenden Maßnahmen in den Lernprozess einzugreifen (Namoun und Alshantiti, 2021, S. 2). Verstärkt wurde dieser Bedarf durch die in der Vergangenheit eingetretenen Covid-Pandemie, und den damit verbundenen Schulschließungen, welche für neue erhebliche Herausforderungen sorgten (Clark et al., 2021, S. 2). Durch (Clark et al., 2021, S. 13) konnte hierzu aufgezeigt werden, welchen positiven Effekt digitale Lernunterstützungen auf die Schülergruppen ausmachen. Aus der Arbeit von (Namoun und Alshantiti, 2021, S. 9) geht jedoch hervor, dass bereits seit 2017 erneut das Interesse anstieg hinsichtlich der Modellierung von Lernergebnissen. Verstärkt wurde dies durch die steigende Verfügbarkeit von digital erfassten Daten durch akademische Informationssysteme und Bildungssoftware (Khan und Ghosh, 2021, S. 1). Seitdem besteht besonders ein Fokus auf die Untersuchung des Bildungsniveaus von Bachelorstudiengängen, sodass die Untersuchung weiterführender Schulen lediglich ein Anteil von in etwa 12% besitzt (Namoun und Alshantiti, 2021, S. 11). Die in den letzten Jahren erforschte Modellierung von Studierendenergebnissen lässt häufig unbeachtet, wie einzelne Faktoren innerhalb der maschinellen Lernmethoden zu den Vorhersagen führen (Namoun und Alshantiti, 2021, S. 19). Die Gesamtheit dieser aktuellen Gegebenheiten motiviert die nachfolgende Untersuchung des gewählten Datensatzes anhand der daran abgeleiteten Forschungsfragen.

2 Zielsetzung

Aus der dargelegten Problemstellung leitet sich für diese Arbeit die folgende Zielsetzung ab:

visuelle Untersuchung der Beziehungen von Einflussfaktoren auf die schulische Leistung im Kontext der Oberschule.

Zielgruppe der Visualisierungen bilden Schülerinnen und Schüler sowie auch Lehr- und Sozialkräfte der schulischen Einrichtungen. Diese tragen u. a. die direkte Lehr- und Integrationsverantwortung, wobei davon auszugehen ist, dass sie durch ein besseres Verständnis der Lerneinflussfaktoren, darin unterstützt werden. Aus dem organisatorischen- und Projektplanungsinteresse kann auch die Schulleitung der Zielgruppe hinzugefügt werden. Angenommen wird, dass die Schulleitung mit der Kenntnis der Einflussfaktoren eine effektivere Planung und Umsetzung von Schulgestaltungsprojekten, durchführen kann.

Trotz des hohen Domänenwissens verfügen damit nur wenige Personen der Zielgruppe über einen Hintergrund in Statistik oder Business Intelligence. Dies ist bei der Entwicklung von Visualisierungen zu berücksichtigen, indem die Komplexität durch bspw. leicht lesbare Diagramme begrenzt wird.

2.1 Datensatz

Der in dieser Arbeit betrachtete Datensatz entstammt der Arbeit von (Cortez und Silva, 2008). Motiviert wurde diese Forschung durch Statistiken, welche Portugal im europäischen Vergleich als deutlich unterdurchschnittlich klassifizierten, aufgrund von hohen Durchfallquoten (Cortez und Silva, 2008, S. 1). Daher wurde mit dieser Arbeit ein realer Datensatz erhoben. Hierfür wurden Schulleisten und schulbezogene Faktoren vom Berichtswesen gesammelt und demografische und soziale Faktoren durch Befragungen ermittelt (Cortez und Silva, 2008, S. 1). Die schulbezogenen Faktoren beziehen sich dabei auf die Leistungen in den Schulfächern Mathematik und Portugiesisch, da Inhalte dieser Fächer übergreifend in anderen Fächern zum Einsatz kommen (Cortez und Silva, 2008, S. 2). Zielgruppe der Untersuchung waren Schüler der dreijährigen zweiten Bildungsphase in Portugal, welche auf der ersten neunjährigen Phase aufbaut (Cortez und Silva, 2008, S. 2). Mittels der Berichte und Umfragen wurden schulbezogene-, demografische-, und soziale Faktoren von 395 Mathematik-Schülern und 649 Portugiesisch-Schülern der *Gabriel Pereira* und der *Mousinho da Silveira* erhoben. Alle erhobenen Faktoren, welche im Rahmen der Analyse auch als Variablen oder Merkmale bezeichnet werden, lassen sich mit ihrem Datentyp der Tabelle 2.1 entnehmen. Detaillierte Informationen zu dessen Erläuterung und ihren Ausprägungen der Tabelle 6.1 im Anhang entnommen werden.

Tabelle 2.1: Kurzbeschreibung der im Datensatz enthaltenen Merkmale

Numerisch	Alter; Mutters Bildungsgrad; Vaters Bildungsgrad; Pendelzeit; Lernzeit; Anzahl bereits durchgefallener Kurse; Qualität der Familienbeziehungen; außerschulische Freizeit; soziale Aktivitäten; Alkoholkonsum an Arbeitstagen; Alkoholkonsum am Wochenende; Gesundheitszustand; Fehltage; erste Vorabnote; zweite Vorabnote; finale Note
Binär	Schulbezeichnung; Geschlecht; Wohngegend; Familiengröße; Zusammenleben der Eltern; externe Lernunterstützung; familiäre Lernunterstützung; bezahlter Extraunterricht; AG-Teilnahme; Besuch der Vorschule; Absicht zur Weiterbildung; häuslicher Internetzugang; Partnerliche Beziehung
Nominal	Mutters Arbeitsbereich; Vaters Arbeitsbereich; Grund der Schulentscheidung; Erziehungsberechtigter

2.2 Forschungsfragen

Folgend gilt es die benannte Zielstellung zu konkretisieren. Dazu werden die für den Rahmen dieser Arbeit zu betrachteten Fragestellungen bestimmt, anhand derer Visualisierungen erstellt und analysiert werden. Beziehen wir hierfür die verschiedenen Handlungsmöglichkeiten der Zielgruppen ein, ergeben sich drei Haupthandlungsfelder für die Beeinflussung der Lernleistung. Diese umfassen die infrastrukturelle Gestaltung der Schule, die individuelle Förderung des sozialen Umfelds der Schüler und die unmittelbare Zusammenarbeit im Unterrichtsformat. Den sich daraus stellenden Forschungsfragen im jeweiligen Handlungsfeld können zusätzlich, nicht vollständig trennscharf, alle Attribute zugewiesen werden.

- Forschungsfrage 1: Welche individuellen Verhaltensweisen und strukturellen Faktoren wirken sich auf die schulischen Leistungen aus ?

Attribute: Geschlecht, Alter, Grund der Schulentscheidung, Pendelzeit, häuslicher Internetzugang, Besuch der Vorschule, Absicht zur Weiterbildung, außerschulische Freizeit, Alkoholkonsum an Arbeitstagen, Alkoholkonsum am Wochenende, Gesundheitszustand

- Forschungsfrage 2: Wie beeinflusst das soziale Umfeld die Schulleistungen der Schülerinnen und Schüler ?

Attribute: Wohngegend, Familiengröße, Zusammenleben der Eltern, Mutters Bildungsgrad, Vaters Bildungsgrad, Mutters Arbeitsbereich, Vaters Arbeitsbereich, Erziehungsberechtigter, Partnerliche Beziehung, Qualität der Familienbeziehungen, soziale Aktivitäten

- Forschungsfrage 3: Wie wird die tatsächliche Bewertung der Lernleistung von der individuellen Leistungsbereitschaft beeinflusst?

Attribute: Lernzeit, Anzahl bereits durchgefallener Kurse, AG-Teilnahme, externe Lernunterstützung, familiäre Lernunterstützung, bezahlter Extraunterricht, Fehltage

3 Grundlagen

Abgeleitet von der zuvor gestellten Forschungsfragen werden in diesem Kapitel notwendige wissenschaftliche Grundlagenkonzepte erläutert. Diese Grundlagenkonzepte entstammen den Forschungsfeldern Data Mining im Bildungsbereich und Business Analytics.

3.1 Data Mining im Bildungsbereich

Data Mining im Bildungsbereich bedeutet, den analytischen Prozess zu beleuchten, welcher Forschenden ermöglicht aus großen Mengen von Rohdaten wertvolles Wissen über bildungspolitische Maßnahmen und Praktiken zu erlangen (Gamazo, 2020, S. 1). Mit der Erforschung dieses Prozesses werden die vier Ziele der Lernverhaltensvorhersage von Schülern, der Erweiterung von Methoden der Psychometrie, das Schaffen von Bildungstemplates und Tools sowie der Ergründung der Wirkung von Lernunterstützung verfolgt (P. Bachhal et al., 2021, S. 2). Das Interesse an diesem Gebiet wird durch die Herausforderung des nicht linearen Einflusses auf die Bildungsleistung durch vielerlei Faktoren geprägt (Khan und Ghosh, 2021, S. 1). Vorhersage, Beziehungsmining und Strukturaufdeckung bilden die bekanntesten Methoden des Data Minings im Bildungsbereich (Gamazo, 2020, S. 3). Anhand von Beziehungsmining werden die stärksten Beziehungen zwischen Merkmalen eines Datensatzes ergründet, ohne vorherige Festlegung von Kriterien oder Prädiktorvariablen (Gamazo, 2020, S. 3).

3.2 Business Analytics

In der Praxis werden Erkenntnisse der Forschung zu Business Analytics dafür eingesetzt, um mittels Organisationsdaten strategische und operationelle Entscheidungen zu fundieren (Qin et al., 2020, S. 1). Diagramme erfüllen dabei den Zweck, ausgerichtet auf Strukturen der visuellen Wahrnehmung, die Identifikation von Mustern zu vereinfachen (Baars und Kemper, 2021, S. 287). Informationswahrnehmung erfolgt durch die Assoziation, Differenzierung und Ordnung von Objekten sowie durch deren Quantität und Größe (Baars und Kemper, 2021, S. 287). Die Gestaltung der Informationswahrnehmung wird zusätzlich unterstützt je konsistenter sich die Darstellung mit dem Vorwissen des Betrachters deckt (Baars und Kemper, 2021, S. 288). Wichtige Schritte im Prozess der Gestaltung umfassen die Manipulation (z. B. Filterung, Aggregation) und das Kartieren (z. B. Zuordnung zu Formen) von Daten (Qin et al., 2020, S. 2). Die Sprache von Visualisierungen besteht somit aus transformierten Daten, in Art, Größe und Farbe unterschiedliche Markierungen und deren Abbildungsverhältnis (Qin et al., 2020, S. 3).

4 Untersuchung der Forschungsfrage

Dieses Kapitel dient der Untersuchung der Forschungsfragen. In Abhängigkeit des untersuchten Handlungsbereichs werden hierfür notwendige Datentransformationen erläutert. Anschließend gilt es aufgrund der Vielzahl an Attributen, die für die Untersuchung relevanten Merkmale zu identifizieren. Final wird die Beantwortung der Forschungsfrage, durch die Vorstellung und Analyse der entwickelten Visualisierungen vorgenommen. Je nach Präsentationsart empfiehlt sich die Darstellung der Visualisierungen anhand exportierter HTML Webseiten oder des im Anhang verfügbaren Jupyter Notebooks. Dies ermöglicht eine Interaktion mit den Visualisierungen durch Funktionen wie dem Hineinzoomen und der Filterung von Kategorien. Im Rahmen dieser Arbeit werden entsprechend dem PDF-Format lediglich statische Abbildungen hinzugefügt.

4.1 Individuelle und strukturelle Faktoren

Zur Untersuchung dieser Faktoren eignet sich besonders ein fachübergreifende Perspektive. Dies resultiert daraus, dass eine unterschiedliche Auswirkung auf die Mathematik- oder Portugiesischlehre nur schwierig kausal zu erklären ist. Die Migrierung beider Datenstände wird anhand der Attribute vorgenommen, welche im, unter (Cortez, [2014](#)) verfügbaren, R-Skript benannt werden. Insgesamt erhält der Datensatz damit die 382 Einträge der Schüler, welche Teil beider Befragungen sind. Sofern unter den ausgewählten Attributen nominale Merkmale vorliegen, welche nicht bei der Datenmigrierung berücksichtigt wurden, kann es hier zu widersprüchlichen Angaben kommen. Diese widersprüchlichen Dateneinträge können nicht ohne weiteres Wissen aufgeklärt, oder ein Mittelmaß bestimmt werden und sind daher zusätzlich aus der Betrachtung zu nehmen.

Die betrachteten Faktoren lassen sich nach individuellen Verhaltensweisen und strukturellen Gegebenheiten aufteilen. Individuelle Merkmale werden nach der Möglichkeit zur Einflussnahme durch den Schüler gefiltert. Weiterhin tritt die Absicht zur Weiterbildung zwar in diesem Kontext als möglicher Prädiktor des Lernerfolgs auf, die alleinige Absicht selbst nimmt jedoch hierauf keinen unmittelbaren Einfluss. Damit werden die Attribute Alkoholkonsum an Arbeitstagen und Alkoholkonsum am Wochenende in die Untersuchung einfließen. Strukturelle Gegebenheiten werden zu dieser Fragestellung durch den Grund der Schulentscheidung, die Pendelzeit und den häuslichen Internetzugang beschrieben. Aus diesen Merkmalen kann besonders der Pendelzeit und dem häuslichen Internetzugang ein potenzieller kausaler Zusammenhang unterstellt werden, anhand des täglichen zusätzlichen Reiseaufwands und dem Zugang zu Onlinewissen. Daher werden diese beiden Attribute für die Untersuchung ausgewählt.

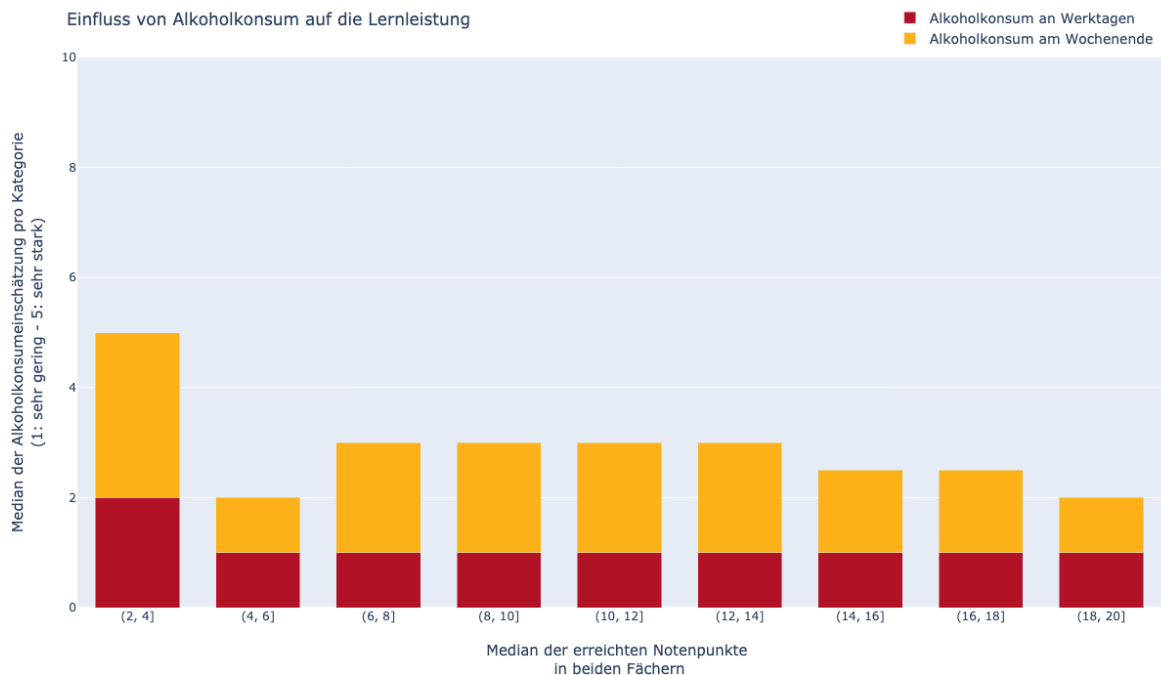


Abbildung 4.1: Einfluss von Alkoholkonsum auf die Lernleistung

Folgend wird der Beantwortung der Forschungsfrage zum Einfluss individueller und struktureller Faktoren auf die Lernleistung eines Schülers nachgegangen. Hierzu dient das in Abbildung 4.1 dargestellte gestapelte Säulendiagramm. Anhand dessen wird als erstes erforscht, wie individuelle Verhaltensweisen auf die Lernleistung wirken anhand des wöchentlichen Alkoholkonsums. Mit dieser Diagrammart wird die Aggregation von Daten notwendig, womit jedoch auch eine Vereinfachung der Verständlichkeit einhergeht. Dem Teil der Zielgruppe ohne statistischen Hintergrund ermöglicht dies besonders zu Beginn einer Präsentation einen leichteren Zugang zu der dargelegten Thematik. Die Grafik ordnet auf der X-Achse die Schüler anhand ihres Medians der finalen Noten beider Fächer in Gruppen ein. Auf der Y-Achse wird der Median des von den Schülern angegebenen Alkoholkonsums auf der Skala von null bis zehn abgetragen. Der Wertebereich ergibt sich aus der Addition der Maximalwerte beider Kategorien. In der Grafik werden die Kategorien des Alkoholkonsums an Arbeitstagen und an Wochenenden berücksichtigt. Farblich sind diese Kategorien zu den Farben von typischen Alkoholprodukten wie Wein und Bier gekennzeichnet, um die Zuordnung zum Thema zu unterstreichen. Zusätzlich können die Farben zu Warnstufen einer Ampel assoziiert werden, bei dem ein Alkoholkonsum an Werktagen häufig kritischer eingeschätzt wird, als an Wochenendtagen. Die Höhe der Säule einer Schülergruppe ergibt sich folglich aus der Summe der Mediane beider Kategorien, um so den gesamten Alkoholkonsum einer Gruppe korrekt darzustellen.

Der Abbildung 4.1 kann entnommen werden, dass die Schülergruppen nahezu alle einen sehr geringen Alkoholkonsum an Werktagen aufweisen, und sich hauptsächlich durch ihren Konsum am Wochenende unterscheiden. Eine Ausnahme bildet dabei die Gruppe, welche im Median beider Fächer die geringste Punktzahl aufweist. Hier liegt bereits der Konsum an Werktagen als einziges leicht über allen anderen Gruppen. Nimmt man die zweite Katego-

rie mit in die Betrachtung auf, zeigt dieselbe Gruppe erneut einen stärkeren Alkoholkonsum auf. Insgesamt liegt die Gruppe mit der geringsten Medianpunktzahl mit 1.5 Einschätzungspunkten über allen anderen Gruppen, was einen negativen Effekt des Alkoholkonsums auf die Lernleistung vermuten lässt. Unterstützt wird diese Vermutung durch die Untersuchung der weiteren Gruppen. Der Verlauf des Alkoholkonsums zeigt hier, dass der Konsum am Wochenende ab einer Mediannote von 14 und 18 Punkten stufenweise absinkt. Eine Ausnahme hierzu stellt die Gruppe im Bereich von vier bis sechs Mediannotenpunkten dar, dessen Konsum gleich dem der Gruppe mit der höchsten Mediannotenpunktzahl ist.

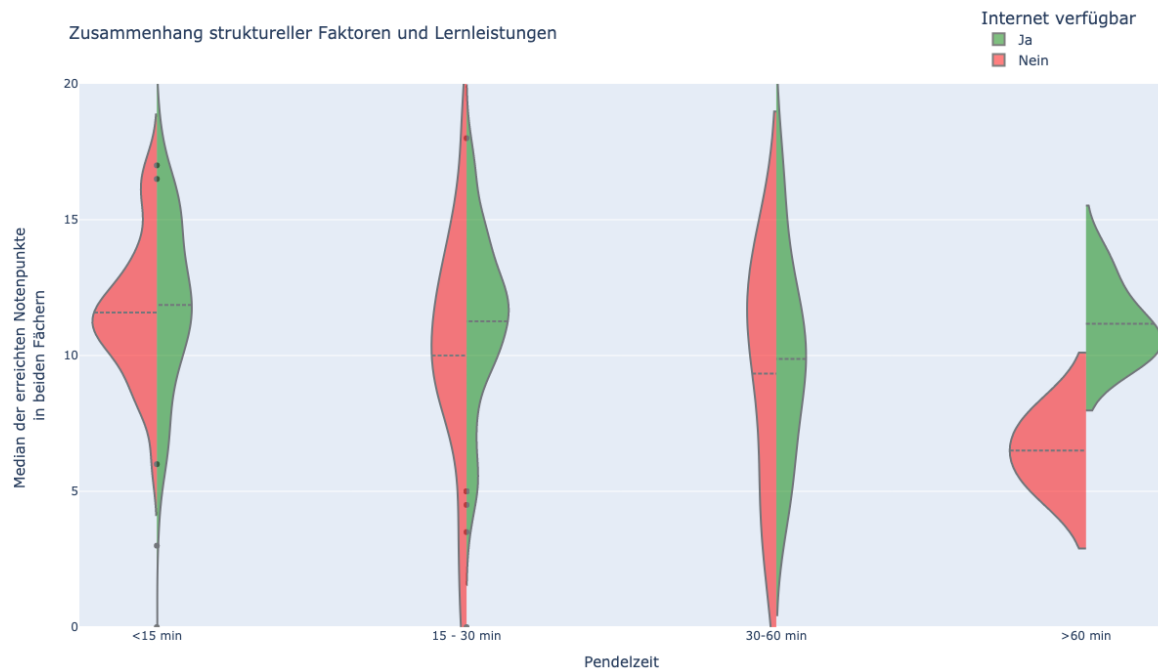


Abbildung 4.2: Zusammenhang struktureller Faktoren und Lernleistungen

Strukturelle Einflüsse auf die Lernleistung werden in dieser Arbeit durch Betrachtung der Pendelzeit und des heimischen Internetzugangs untersucht. Für die Untersuchung dient der in Abbildung 4.2 dargelegte Violinen-Plot. Dargestellt wird mit dieser Diagrammart die Datenverteilung der im Median erreichten Notenpunktzahl in Portugiesisch und Mathematik, im Bezug auf die gewählten Variablen. Hierbei legt die X-Achse die Pendelzeit anhand ihrer vier möglichen Werte ($<15 \text{ min}$, $15-30 \text{ min}$, $30-60 \text{ min}$, $>60 \text{ min}$) und die Y-Achse die im Median erreichte Notenpunktzahl beider Fächer dar. Eine *Violine* wird dabei in der Verteilung linksseitig durch die der Gruppe des vorhandenen Internetzugangs und rechtsseitig durch die Gruppe des fehlenden Internetzugangs bestimmt. Zu jeder Verteilung sind dessen Median als gestrichelte Linie und Ausreißer als Punkte erkennbar. Die Farbe einer Verteilung orientiert sich hier anhand der Kategorie und folgt dem allgemeinen Farbverständnis von Grün als Bestätigung und Rot als Ablehnung des Internetzugangs.

Mittels der Untersuchung der Abbildung 4.2 und dessen Verlauf über alle Pendelzeiten wird zunächst deutlich, dass die gezeigten Mittelwerte mit steigender Pendelzeit tendenziell abnehmen. Wird hinzu die Verteilung der Lernleistung betrachtet, ist dieser Trend aufgrund der hohen Quartilsabstände teilweise zu relativieren. Einzige Ausnahme dieses Abnahmetrends

zeigt die Schülergruppe, welche eine Pendelzeit von mehr als 60 Minuten besitzt sowie über heimischen Internetzugang verfügt. Besonders deutlich in dieser Pendelzeitgruppe wird zudem der im Verhältnis hohe Abstand von 4.25 Notenpunkten zwischen den Mittelwerten in Abhängigkeit des Internetzugangs. Betont wird dies zugleich durch die deutlich geringeren Quartilsabstände der Notenverteilung in dieser Pendelzeitgruppe. Da ohne weiteres Wissen lediglich Vermutungen über die kausalen Gründe dieses Phänomens aufgestellt werden können, gilt es dieses in weiteren Arbeiten detaillierter zu Erforschen.

4.2 Soziales Umfeld

Der Einfluss des sozialen Umfelds eines Schülers auf dessen Lernleistung wird gleich der vorherigen Sektion aus fachübergreifender Perspektive betrachtet. Dadurch kommen für die folgende Untersuchung die gleichen Datentransformationen zur Trage.

In der Auswahl von Attributen wird sich zum einen auf das familiäre Umfeld beschränkt. Dabei werden konkret die Merkmale Familiengröße und Qualität der Familienbeziehungen untersucht und die Information zum Zusammenleben der Eltern, deren Bildungsgrad und Arbeitsbereich und zum Sorgerecht exkludiert. Diese Entscheidung begründet sich in der Annahme, dass sich die Effekte der ausgeschlossenen Variablen in der Qualität der Familienbeziehungen konsolidiert widerspiegeln. Zum anderen werden die Variablen Partnerstatus und Intensität sozialer Aktivitäten der Analyse hinzugefügt, um das soziale Verhältnis mit Gleichaltrigen zu beleuchten.

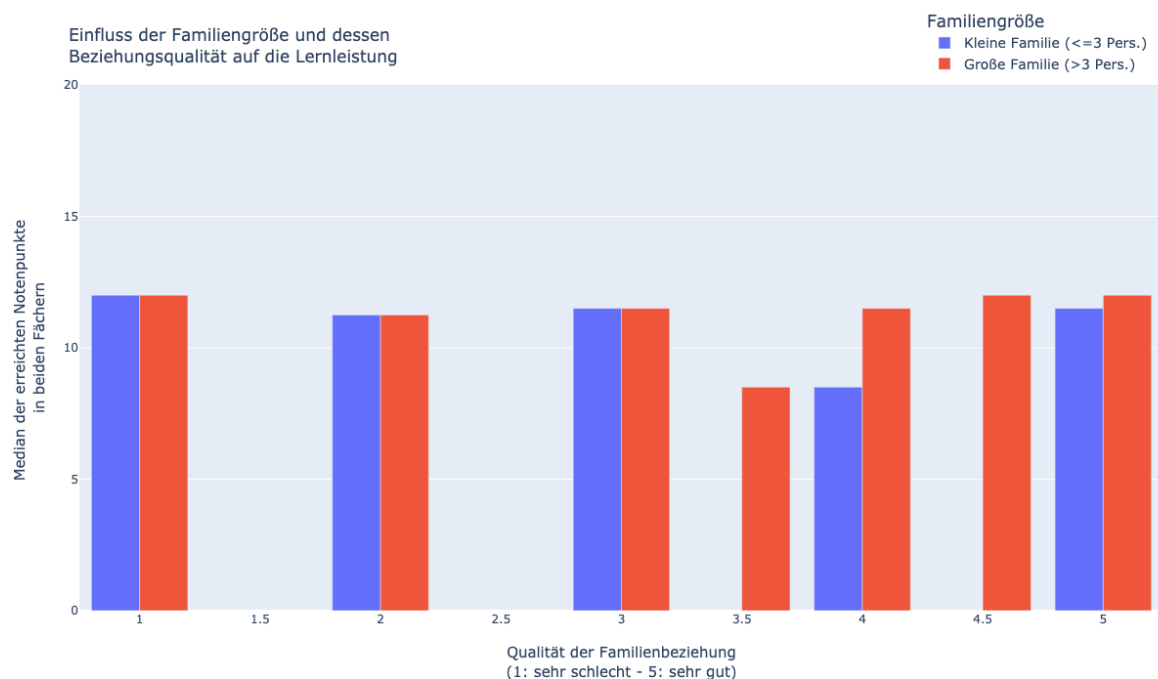


Abbildung 4.3: Einfluss der Familiengröße und dessen Beziehungsqualität auf die Lernleistung

Abbildung 4.3 zeigt ein Säulendiagramm, welches die Beziehung zwischen Familiengröße, Beziehungsqualität und der Lernleistung thematisiert. Daran wird folglich die Forschungsfrage nach dem Einfluss des sozialen Umfelds auf die Lernleistung genauer beleuchtet. Lern-

leistung wird als Median der erreichten Notenpunktzahl beider Fächer auf der Y-Achse im vollen möglichen Bereich von null bis 20 dargestellt. Dabei wird der allgemeine Konsens angenommen, dass potenziell abhängige Variablen auf dieser Y-Achse dargelegt werden, um der Gewohnheit der Zielgruppe zu entsprechen. Die X-Achse erfüllt zugleich den Zweck, den Median der Qualität der Familienbeziehungen abzubilden. Durch die Aggregation der Daten aus den zwei Befragungen entstehen hieraus auch ungerade Werte und damit einzelne Säulen der Beziehungsqualität in lediglich einer Familiengrößenkategorie. Kategorien der Familiengröße werden zu jeder Beziehungsqualitätsgruppe als Säulen, farblich möglichst kontrastreich nebeneinander platziert, was dem Betrachter den Vergleich dieser Werte vereinfacht.

Die Analyse der Abbildung 4.3 zeigt über alle Qualitäten der Familienbeziehungen nahezu die gleichen Werte der im Median erreichten Notenpunkte. Aus der Betrachtung dieses Datensatzes kann somit kein Einfluss der Güte der Familienbeziehung auf die Lernleistung in Form von besseren Schulnoten vermutet werden. Hingegen ist der Grafik das Phänomen zu entnehmen, dass zumindest unter besseren Familienbeziehungen und einer größeren Anzahl von Familienmitgliedern auch eine geringe höhere Lernleistung auftritt. Zweck weiterer Arbeiten könnte es damit sein, dieses wie auch bisherige Phänomene anhand neuer Analysen tiefergehend zu beleuchten.

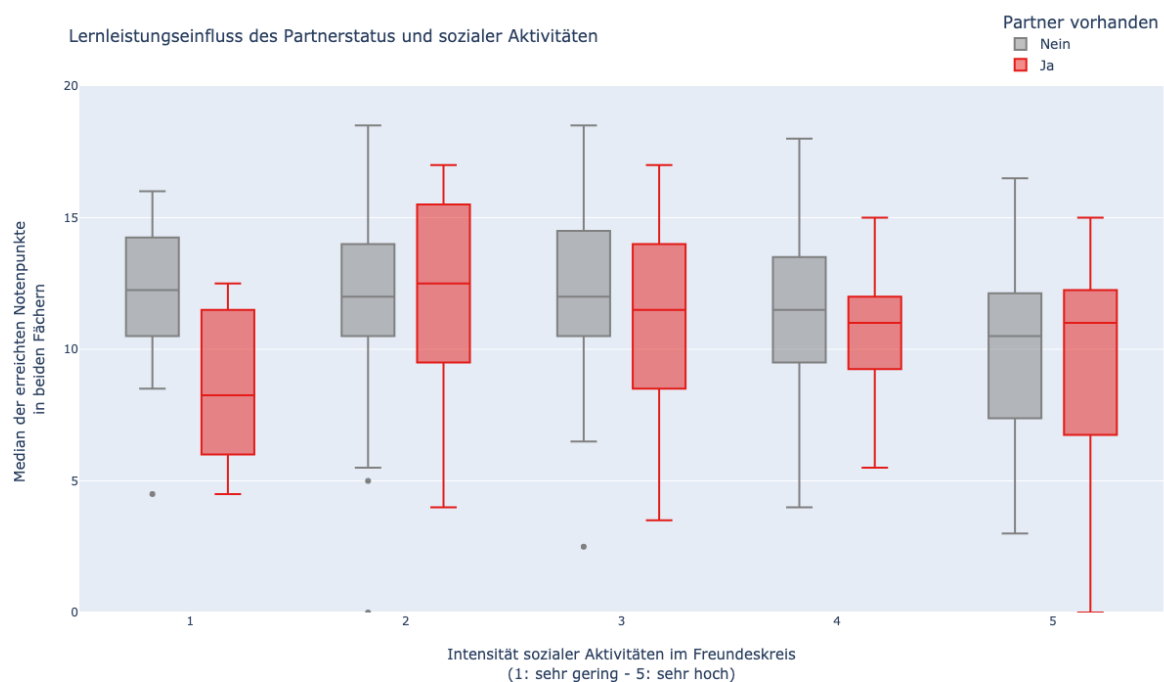


Abbildung 4.4: Lernleistungseinfluss des Partnerstatus und sozialer Aktivitäten

Die Untersuchung des freundschafts- und partnereinflusses auf die Lernleistung wird anhand der Abbildung 4.4 durchgeführt. Darin wird zu jedem Wert der unabhängigen Variablen die Verteilung der abhängigen Variable als Boxplot abgebildet. Entsprechend der häufigen Darstellung werden die unabhängigen Variablen auf der X-Achse dargestellt. Dies erfolgt zum einen durch das Aufzeigen aller möglichen Werte (eins bis fünf) der Intensität sozialer Aktivitäten. Zum anderen werden jedem Intensitätsgrad zwei Boxplots entsprechend der möglichen Partnerstatus zugewiesen. Ein Partnerstatus wird dabei farblich kodiert, sodass

das Vorhandensein eines Partners entsprechend der damit einhergehenden Assoziation mit Liebe, zu einem tiefroten Boxplot führt. Das Fehlen eines Partners wird durch das soziale Grundbedürfnis eines Menschen als tendenziell einsamer und weniger aufregend eingeschätzt und damit durch die graue Farbe ausgedrückt.

Folgend wird die Korrelation der abhängigen Variable (Median der erreichten Note) und der unabhängigen Variablen (Intensität sozialer Aktivitäten, Partnerstatus) untersucht. Der Erkenntnisgewinn korrelierender Effekte gilt es in weiteren Arbeiten auf kausalen Zusammenhang zu prüfen. Ein solcher korrelierender Effekt kann die bei der Betrachtung der Intensität sozialer Aktivitäten vermutet werden. Hierbei zeigt sich, bis auf bei der Gruppe mit einer sehr geringen Aktivität, ein sehr leicht abnehmender Trend der im Median erreichten Note, unter steigender Intensität an Sozialaktivitäten. Weiterhin auffällig ist dazu die Inspektion des Partnerstatus bei sehr geringer sozialer Aktivität. In dieser Gruppe von Schülern zeigt sich im Median eine Verringerung der Notenpunktzahl um vier Punkte unter denen, welche eine partnerliche Beziehung führen. Kausale Vermutungen, wie das eine partnerliche Beziehung unter zugleich geringer anderweitiger sozialer Aktivitäten zu mehr Streit und damit geringerer Lernleistung führt, bedarf einer Validierung durch Analysen weiterer Arbeiten.

4.3 Individuelle Leistungsbereitschaft

In der Untersuchung des Einflusses individueller Leistungsbereitschaft auf die Lernleistung werden die ursprünglichen fachspezifischen Datensätze verwendet. Grund dafür ist der fehlende kausale Zusammenhang zwischen der Leistungsbereitschaft in einem Fach und der erzielten Lernleistung im anderen Fach. Zu Berücksichtigen sind damit die unterschiedliche Datensatzgröße mit 395 (Mathematik-Schüler) und 649 (Portugiesisch-Schüler) Einträgen. Entsprechend werden in dieser Sektion alle entwickelten Visualisierungen jeweils mit beiden Datengrundlagen dargestellt.

Von den, der Forschungsfrage zugeordneten Attributen, zeigen Lernzeit, Anzahl bereits durchgefallener Kurse, externe Lernunterstützung, familiäre Lernunterstützung, bezahlter Extrunterricht und Fehltage eine besondere Relevanz für die Lernleistung. In dieser Ergründung der Einflussfaktoren wird die AG-Teilnahme exkludiert, da hiermit zwar Zusatzarbeiten, jedoch abseits vom fachlichen Thema, geleistet werden. Zusätzlich wird die Anzahl bereits durchgefallener Kurse nicht beachtet, weil hierbei kein kausaler Zusammenhang zur aktuellen Lernleistung besteht.

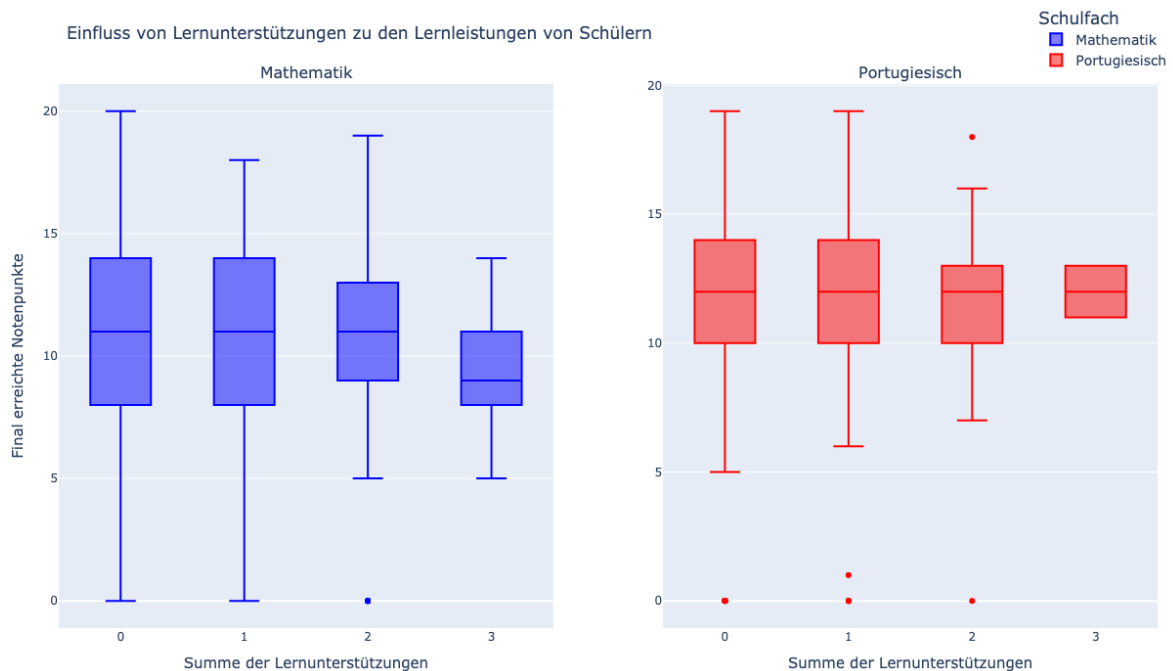


Abbildung 4.5: Einfluss von Lernunterstützungen zu den Lernleistungen von Schülern

Mit Abbildung 4.5 wird dem Einfluss der individuellen Leistungsbereitschaft auf die Lernleistung nachgegangen. Wie zu Beginn dieser Sektion erläutert, wird keine fachübergreifende Perspektive eingenommen, weshalb folglich zwei Grafiken dargestellt sind. Mittels der Grafiken werden visuell die Abhängigkeitsbeziehung der Lernleistung von familiär, schulisch oder bezahlt erhaltenen Lernunterstützungen ergründet. Eine Grafik enthält ein Boxplot für jeden Wert der abhängigen Variable, welcher auf der Y-Achse die zugehörige Verteilung der erreichten Note anzeigt. Die visuelle Skala der Y-Achse entspricht dabei der möglichen Punkteskala einer Note. Durch die zweierlei gleiche Skala der Y-Achse kann dem Betrachter der Zielgruppe zugemutet werden zu erkennen, dass hier lediglich eine Y-Achsenbeschriftung für beide Grafiken gilt. Damit kann der Abbildung Text entnommen, und so diese übersichtlicher gestaltet werden. Auf der X-Achse wird die Summe der von einem Schüler erhaltenen Lernunterstützung, entsprechend in der Skala von null bis drei abgebildet. Die Boxplots beider Grafiken sollen vom Betrachter durch ihre Farbgebung möglichst leicht den einzelnen Schulfächern zugeordnet werden können. Unterstützt wird dies damit, dass die Grafik der Befragung im Fach Portugiesisch farblich mit dem Rot der portugiesischen Flagge gekennzeichnet ist (Wikipedia, 2024). Dazu wurde die Grafik der Befragung im Fach Mathematik möglichst kontrastreich in Blau markiert.

Die Analyse der Abbildung 4.5 zeigt zwei hervorzuhebende Besonderheiten auf. Anhand der Beobachtung des Mittelwerts kann im Fall beider Lernleistungen vermutet werden, dass allgemein mit höherem Erhalt von Lernunterstützung, die Lernleistung wider erwartend abnimmt oder maximal gleich bleibt. Weiterhin wird durch die Quartilsabstände vermutbar, dass Lernunterstützungen sogar fachübergreifend die Streuung der Lernleistung deutlich verringern können.

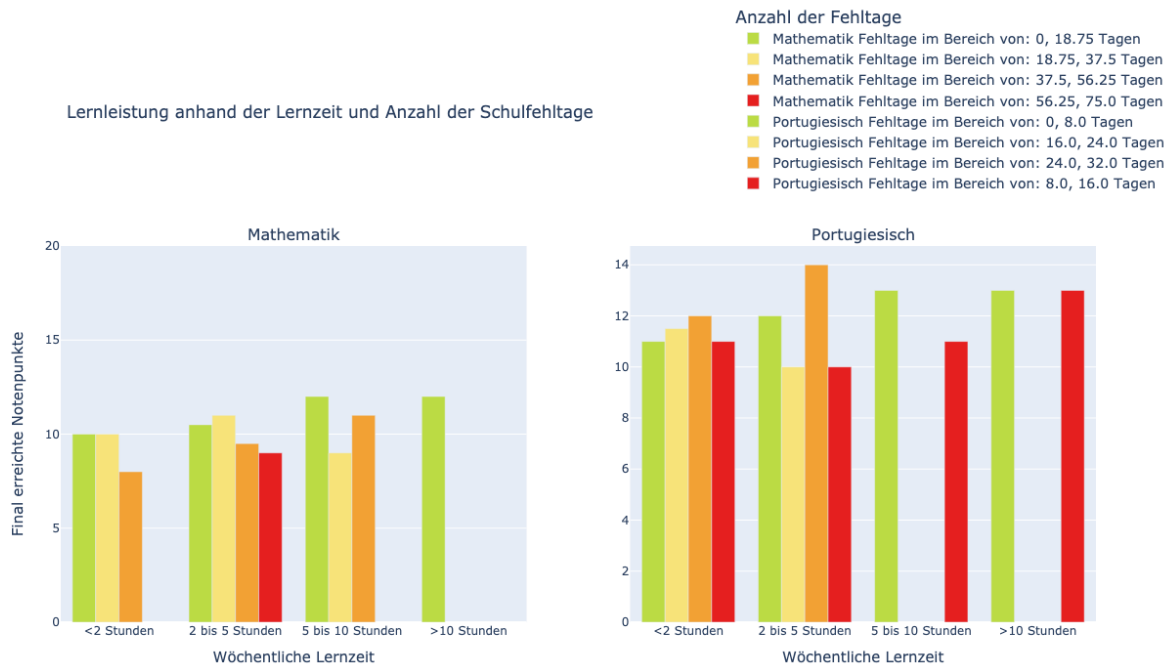


Abbildung 4.6: Lernleistung anhand der Lernzeit und Anzahl der Schulfehltag

Ein weiterer Aspekt der dritten Forschungsfrage wird durch Abbildung 4.6 analysiert. Beide Grafiken zeigen die Beziehung auf der Lernzeit und der Menge der Fehltage zu der erreichten Note. Die X-Achse einer Grafik drückt dabei zum einen die vier Werte der wöchentlichen Lernzeit aus. Über mehrere Säulen wird die Menge der Fehltage in vier Bereiche kategorisiert. Die Farben der Säulen sind im Verlauf von Grün zu Rot dargestellt, um die Gefährdung durch Wissensbenachteiligung zu signalisieren.

Über beide Grafiken hinweg lässt sich erkennen, dass in mehreren Kategorien eine Erhöhung der Lernzeit zu einer gesteigerten Lernleistung führt. Die Lernleistung, widerspiegelt in der Note, folgt undeutlicher dem Trend, dass eine zunehmende Abwesenheit geringere Noten hervorruft. Dieser Effekt tritt nur im Mathematikbereich bis zu einer Lernzeit von fünf Stunden auf. Hingegen klarer erkenntlich ist, dass unabhängig der betrachteten Variablen die Noten in Mathematik im Median geringer sind, als in Portugiesisch.

5 Fazit und Ausblick

Abschließend werden anhand dieses Kapitels die Ergebnisse der Untersuchungen zusammengefasst. In dieser Arbeit wurde sich der Zielstellung gewidmet, die Beziehungen von Einflussfaktoren auf die schulische Leistung im Kontext der Oberschule zu untersuchen. Dies wurde durchgeführt anhand der Forschungsfragen zur individueller Lernleistung, zum sozialem Umfeld und zu individuellen Verhaltensweisen und strukturellen Faktoren. Individuelle Verhaltensweisen wie der Konsum von Alkohol zeigten dabei einen negativ korrelierenden Effekt. Ebenfalls wies der Datensatz eine Vermutung zu einem negativen Effekt steigender Pendelzeit und einem positiven Effekt von Internetzugang auf. Das soziale Umfeld beeinflusste die Schüler des Datensatzes zum einen positiv mittels der Größe von intensiven Familienbeziehungen. Zum anderen zeigte die Intensivierung sozialer Aktivitäten auf eine negativ korrelierende Lernleistung. Die individuelle Leistungsbereitschaft wies einen unterstützenden Trend der Noten durch steigern der Lernzeit auf. Hingegen beschränkte sich der Effekt von Lernunterstützungen auf das Verdichten der Notenergebnisse.

Alle erläuterten Korrelationen werden in ihrer Aussagekraft limitiert durch die bewusste Beschränkung auf die final erreichte Note als Bemessung der Lernleistung. Weiterhin zeigt das Forschungsfeld des Data Minings im Bildungswesen eine Vielzahl weiterer erhobener Daten auf. Insgesamt führt dies dazu, dass es notwendig ist, alle in dieser Arbeit erkannten Phänomene und Korrelationen mittels zusätzlicher Daten zu validieren und auf Kausalität zu prüfen. So kann erzielt werden, dass durch neue Erkenntnisse alle Zielgruppen in ihrer Aufgabe im Bildungssystem unterstützt und damit das Bildungsniveau angehoben werden kann.

Literatur

- Baars, H., & Kemper, H.-G. (2021). Informationsbereitstellung. In *Business Intelligence & Analytics – Grundlagen und praktische Anwendungen: Ansätze der IT-basierten Entscheidungsunterstützung* (S. 287–321). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-8348-2344-1_4
- Clark, A. E., Nong, H., Zhu, H., & Zhu, R. (2021). Compensating for academic loss: Online learning and student performance during the COVID-19 pandemic. *China Economic Review*, 68, 101629. <https://doi.org/10.1016/j.chieco.2021.101629>
- Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. <https://api.semanticscholar.org/CorpusID:16621299>
- Cortez, P. (2014). Student Performance [DOI: <https://doi.org/10.24432/C5TG7T>]. <https://archive.ics.uci.edu/dataset/320/student+performance>
- Gamazo, F., Adriana and Martínez-Abad. (2020). An Exploration of Factors Linked to Academic Performance in PISA 2018 Through Data Mining Techniques. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.575167>
- Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and Information Technologies*, 26(1), 205–240. <https://doi.org/10.1007/s10639-020-10230-3>
- Namoun, A., & Alshanqiti, A. (2021). Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. *Applied Sciences*, 11(1). <https://doi.org/10.3390/app11010237>
- P. Bachhal, S. Ahuja & S. Gargish. (2021). Educational Data Mining: A Review. *Journal of Physics: Conference Series*, 1950(1), 012022. <https://doi.org/10.1088/1742-6596/1950/1/012022>
- Qin, X., Luo, Y., Tang, N., & Li, G. (2020). Making data visualization more efficient and effective: a survey. *The VLDB Journal*, 29(1), 93–117. <https://doi.org/10.1007/s00778-019-00588-3>
- Wikipedia. (2024). National colours of Portugal [zuletzt aufgerufen am 20. März 2024]. https://en.wikipedia.org/w/index.php?title=National_colours_of_Portugal&oldid=1205129773

6 Anhang

Tabelle 6.1: Detaillierte Darstellung des untersuchten Datensatzes

Beschreibung	Datentyp	Ausprägungen
Schulbezeichnung	Binär	GP - Gabriel Pereira; MS - Mousinho da Silveira
Geschlecht	Binär	F - weiblich; M - männlich
Alter	Numerisch	15 bis 22
Wohngegend	Binär	U - urban; R - ländlich
Familiengröße	Binär	LE3 - kleiner oder gleich 3; GT3 - mehr als 3
Zusammenleben der Eltern	Binär	T - leben gemeinsam; A - leben auseinander
Mutters Bildungsgrad	Numerisch	0 - kein Bildungsgrad; 1 - Grundschulabschluss (4. Klasse); 2 - 5. bis 9. Klasse; 3 - Oberstufenabschluss; 4 - höherer Bildungsgrad
Vaters Bildungsgrad	Numerisch	0 - kein Bildungsgrad; 1 - Grundschulabschluss (4. Klasse); 2 - 5. bis 9. Klasse; 3 - Oberstufenabschluss; 4 - höherer Bildungsgrad
Mutters Arbeitsbereich	Nominal	Lehrerin; Gesundheitswesen; Sozialwesen (Verwaltung oder Polizei); zu Hause; anderer Bereich
Vaters Arbeitsbereich	Nominal	Lehrer; Gesundheitswesen; Sozialwesen (Verwaltung oder Polizei); zu Hause; anderer Bereich
Grund der Schulentscheidung	Nominal	Nahe dem Zuhause; Ruf der Schule; Kurspräferenz; anderer Grund
Erziehungsberechtigter	Nominal	Mutter; Vater; Anderer
Pendelzeit	Numerisch	1 - <15 min.; 2 - 15 bis 30 min.; 3 - 30 min. bis 1 Stunde; 4 - >1 Stunde
Lernzeit	Numerisch	1 - <2 Stunden; 2 - 2 bis 5 Stunden; 3 - 5 bis 10 Stunden; 4 - >10 Stunden
Anzahl bereits durchgefallener Kurse	Numerisch	1 bis 3; 4
externe Lernunterstützung	Binär	Ja; Nein
familiäre Lernunterstützung	Binär	Ja; Nein
bezahlter Extraunterricht	Binär	Ja; Nein
AG-Teilnahme	Binär	Ja; Nein
Besuch der Vorschule	Binär	Ja; Nein
Absicht zur Weiterbildung	Binär	Ja; Nein
häuslicher Internetzugang	Binär	Ja; Nein
Partnerliche Beziehung	Binär	Ja; Nein
Qualität der Familienbeziehungen	Numerisch	von 1 - sehr schlecht bis 5 - exzellent
außerschulische Freizeit	Numerisch	von 1 - sehr schlecht bis 5 - sehr gut
soziale Aktivitäten	Numerisch	von 1 - sehr schlecht bis 5 - sehr gut
Alkoholkonsum an Arbeitstagen	Numerisch	von 1 - sehr schlecht bis 5 - sehr gut
Alkoholkonsum am Wochenende	Numerisch	von 1 - sehr schlecht bis 5 - sehr gut
Gesundheitszustand	Numerisch	von 1 - sehr schlecht bis 5 - sehr gut
Fehltage	Numerisch	von 0 bis 93
erste Vorabnote	Numerisch	von 0 bis 20
zweite Vorabnote	Numerisch	von 0 bis 20
finale Note	Numerisch	von 0 bis 20

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit eigenständig und ohne fremde Hilfe angefertigt habe. Textpassagen, die wörtlich oder dem Sinn nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche kenntlich gemacht.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Köln, den 22. März 2024

Leon Henne