

Hausarbeit

Einflussfaktoren auf die Schulleistungen im Rahmen der Oberschule

MADS2100 Reporting und Visualisierung 23oB

Leon Henne

Köln, den 14. 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 how to Visualisierungen ?	4
3.2 how to analyse Visualisierungen ?	4
4 Untersuchung der Forschungsfrage	5
4.1 Individuelle und strukturelle Faktoren	5
4.2 Soziales Umfeld	5
4.3 Individuelle Leistungsbereitschaft	6
5 Fazit	7
Literatur	8

Tabellenverzeichnis

2.1	Kurzbeschreibung der im Datensatz enthaltenen Merkmale	3
5.1	Detaillierte Darstellung des untersuchten Datensatzes	10

Abbildungsverzeichnis

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 Alshanqiti, 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 Alshanqiti, 2021, S. 9) geht jedoch hervor, dass bereits seit 2017 erneut das Interesse anstieg hinsichtlich der Modellierung von Lernergebnissen. 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 Alshanqiti, 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 Alshanqiti, 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 5.1 im Anhang entnommen werden.

Tabelle 2.1: Kurzbeschreibung der im Datensatz enthaltenen Merkmale

Numerisch	Alter; Mutters Bildungsgrad; Vaters Bildungsgrad; Pendelzeit; Lernzeit; Anzahl durchgefallener Kurse; Qualität der Familienbeziehungen; außerschulische Freizeit; soziale Aktivitäten; Alkoholkonsum unter der Woche; 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 Unterrichtsfeld. Den sich daraus stellenden Forschungsfragen im jeweiligen Handlungsfeld können zusätzlich, nicht vollständig trennscharf, die Attribute zugewiesen werden.

- Forschungsfrage 1: Welche individuellen 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 unter der Woche, 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 durchgefallener Kurse, AG-Teilnahme, externe Lernunterstützung, familiäre Lernunterstützung, bezahlter Extraunterricht, Fehltage

3 Grundlagen

3.1 how to Visualisierungen ?

3.2 how to analyse Visualisierungen ?

4 Untersuchung der Forschungsfrage

Dieses Kapitel dient der Untersuchung der Forschungsfragen. In Abhängigkeit des untersuchten Handlungsbereichs werden die hierfür notwendigen 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.

4.1 Individuelle und strukturelle Faktoren

Zur Untersuchung dieser Faktoren eignet sich besonders ein fachübergreifende Perspektive, da eine fachspezifische unterschiedliche Auswirkung nur schwierig kausal zu erklären ist. Im Weiteren wird dadurch eine Erhöhung der Datenmenge erzielt, da alle Attribute vollständig in dem Datensatz der Portugiesisch-Schüler und Mathematik-Schüler enthalten sind. Die Migrierung beider Datenstände wird anhand der Attribute vorgenommen, welche im, unter (Cortez, 2014) verfügbaren, R-Skript benannt werden. Insgesamt erhält der Datensatz damit 657 Einträge.

Mit der in dieser Sektion betrachteten Faktoren ergibt sich die gleichnamige Aufteilung der Attribute zu den Aspekten der individuellen Eigenschaften und Verhaltensweisen und der strukturellen Gegebenheiten. Individuelle Merkmale lassen sich zum einen durch die Möglichkeit zur aktuellen Einflussnahme durch den Schüler selbst filtern. Zum Anderen tritt die Absicht zur Weiterbildung zwar in diesem Kontext als Prädiktor des Lernerfolgs auf, die alleinige Absicht selbst nimmt jedoch hierauf keinen unmittelbaren Einfluss. Damit werden die Attribute *außerschulische Freizeit*, *Alkoholkonsum unter der Woche*, *Alkoholkonsum am Wochenende*, *Gesundheitszustand* in die Untersuchung einfließen. Strukturelle Gegebenheiten werden im Kontext dieser Arbeit 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 anhand des täglichen zusätzlichen Reiseaufwands und dem Zugang zu Onlinewissen unterstellt werden. Daher werden dazu diese beiden Attribute für die Untersuchung ausgewählt.

4.2 Soziales Umfeld

- Erläuterung der Datenstruktur: merged, left join, individual (mat_df, por_df)
- Auswahl der Variablen in Abhängigkeit zur Frage:
address, famsize, Pstatus, medu, fedu, mjob, fjob, guardian, romantic, famrel, goout
- Grafik idee:

Scatterplot: Bestimmen von Personengruppen anhand von visuellem (1-2) oder maschinellem (3-x) Clustering.

Säulendiagramm: Median der Notenleistung (G3) zu den einzelnen Personengruppen

4.3 Individuelle Leistungsbereitschaft

- Erläuterung der Datenstruktur: merged, left join, individual (mat_df, por_df)
- Auswahl der Variablen in Abhängigkeit zur Frage:
studytime, failures, activities, schoolsup, famsup, paid, absences
- Grafik idee:

5 Fazit

Literatur

- 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>
- 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>

Tabelle 5.1: 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 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 unter der Woche	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 14. März 2024

Leon Henne