

Ruhr-Universität Bochum
Lehrstuhl für Data Science in Economics
Prof. Dr. Astrid Krenz
Email: astrid.krenz@ruhr-uni-bochum.de

Bachelorseminar in Data Science

Wintersemester 2023 / 2024

Dozentin

Prof. Dr. Astrid Krenz

Kursinformation

Das Seminar ist anrechenbar mit 5 Credit Points unter Modulnummer 073150. Die maximale Anzahl an Teilnehmer*innen beträgt 20.

Lernziele

Das Seminar richtet sich an Bachelor-Studierende und behandelt Themen aus dem Bereich der Data Science. Die Teilnehmer*innen erhalten einen Überblick über Entwicklungen in diesem Forschungsfeld. Sie lernen die relevante Literatur zu analysieren und zu diskutieren. Die Studierenden üben Fähigkeiten im wissenschaftlichen Schreiben und Präsentationstechniken ein.

Kursinhalte

Das Seminar behandelt verschiedene, aktuelle Themen aus dem Bereich der Data Science, wie z. B. Classification, Clustering, Dimensionality Reduction, NLP und deren Anwendungen in den Wirtschaftswissenschaften, unter anderem in Monetary Economics, International Economics, Labour Economics, Regional Economics, Environmental Economics, Development Economics.

Zeit und Ort

Das Seminar findet geblockt am 30.1.2024 von 13:00 bis 18:00 Uhr und am 31. Januar 2024 von 9:00 bis 18:00 Uhr statt. Eine Einführungsveranstaltung findet am 17. Oktober 2023 um 14:15 Uhr statt. Die Teilnehmer*innen werden per Email über die Räume für die Treffen informiert. Bitte kontrollieren Sie Ihren Email-Account entsprechend regelmäßig.

Bewerbung und Registrierung in FlexNow

Sie können sich bis zum 9. Oktober 2023 um einen Seminarplatz bewerben, indem Sie sich in Moodle eintragen und eine Email an die Email-Adresse astrid.krenz@ruhr-uni-bochum.de mit den folgenden Informationen senden:

1. Ihren Namen, Ihre Matrikelnummer, Ihr Studienfach, Ihr Studiensemester und Ihre aktuelle Email-Adresse
2. den Erst-, Zweit- und Drittwunsch für ein Seminarthema (mit Angabe der Nummern 1. bis 10.)
3. eine Angabe, an welchen anderen Bachelor-Seminaren Sie teilgenommen haben (und einen Beleg dazu mit beifügen)
4. welche Kurse in Statistik, Ökonometrie und Empirischer Wirtschaftsforschung Sie bislang belegt haben und die Noten, die Sie in diesen Kursen erhalten haben (und einen Beleg dazu mit beifügen)
5. eine Angabe, ob Sie an der Ruhr-Universität Bochum, an einer anderen UA-Ruhr Universität oder im Erasmus-Programm (bitte auch den Namen der Heimathochschule angeben) studieren.

Als Beleg für die in den Punkten 3. und 4. genannten Angaben können Sie eine Notenübersicht mit beilegen, in der alle nicht abgefragten anderen Informationen unkenntlich gemacht sind.

Bitte beachten Sie, dass für die Auswahl nur die Bewerbungen berücksichtigt werden können, die Informationen zu allen fünf Punkten enthalten. Sie erhalten von mir bis zum 11. Oktober 2023 eine Email, in der ich Sie darüber informiere, ob Ihre Bewerbung erfolgreich war.

Um Credit-Points für das Seminar zu erhalten, ist eine Anmeldung im Prüfungssystem FlexNow erforderlich. Die Anmelde- bzw. Abmeldefrist in FlexNow endet am 27. Oktober 2023.

Seminararbeit

Die schriftliche Seminararbeit muss bis spätestens zum 9. Januar 2024 um 12 Uhr mittags eingereicht werden unter der Email-Adresse astrid.krenz@ruhr-uni-bochum.de, entweder als pdf- oder doc-Dokument. Detaillierte Informationen dazu, was beim Schreiben einer Seminararbeit und bei der Präsentation zu beachten ist, werden während der Einführungsveranstaltung am 17. Oktober 2023 vermittelt. Nach der Einführungsveranstaltung werden wir individuelle Treffen miteinander vereinbaren, um z.B. über die Gliederung der Seminararbeit und den Fokus Ihrer Arbeit zu sprechen.

Vorkenntnisse

Obwohl keine Vorkenntnisse für das Seminar vorausgesetzt werden, werden gute Kenntnisse in Grundlagen in Data Science, Statistik und Ökonometrie empfohlen.

Richtlinien

Die Credit-Punkte werden für die Ausarbeitung einer schriftlichen Seminararbeit vergeben. Zudem ist eine Präsentation über die eigene Seminararbeit zu halten und eine Teilnahme an der Diskussion zu den anderen Seminarthemen erforderlich. In der Seminararbeit zeigen die Studierenden, dass sie die relevante Literatur sichten, analysieren und wichtige Ergebnisse daraus extrahieren können. Die Studierenden demonstrieren ihre Fähigkeiten im wissenschaftlichen Schreiben und motivieren in ihrer Seminararbeit ein Thema, erläutern in strukturierter Weise die Methoden und Argumente in der Literatur und setzen sich kritisch mit ihnen auseinander. Auf der Grundlage des ihnen zugeteilten topics und des research papers (siehe list of topics und research papers unten) sollen die Studierenden die relevanten angewendeten Methoden aus der Data Science beschreiben, den Beitrag der Nutzung dieser Methoden aus der Data Science erklären und diesen mit Ergebnissen aus der Forschungsliteratur, die mit anderen/ konventionellen Forschungsmethoden arbeitet, vergleichen. In Absprache koennen fuer die Seminararbeit auch eine eigene Datenarbeit und Programmierungen vorgenommen werden, im Rahmen des vorgegebenen topics/ research papers. Die Seminararbeit soll 10 Seiten (reiner Textteil) umfassen. Bei der Präsentation demonstrieren die Studierenden, dass sie komplexe Theoriemodelle und empirische Studien erfassen und kompakt wiedergeben können. Die Vortragszeit für die Präsentation beträgt 20 Minuten.

Für alle Sitzungen gilt Anwesenheitspflicht.

Ihre Seminararbeit wird mit einer speziellen Plagiatssoftware untersucht. Beachten Sie, korrekt zu zitieren, da das Auffinden eines Plagiats dazu führt, dass Ihnen für das Seminar die Note 5,0 gegeben wird.

Benotung

Für die schriftliche Seminararbeit gibt es eine Note. Um den Kurs zu bestehen, ist das Halten einer Präsentation und eine Teilnahme an der Diskussion im Seminar erforderlich. Sowohl der schriftliche als auch der mündliche Teil müssen mit mindestens der Note 4,0 (bestanden) bewertet worden sein, um die Credit-Points für das Seminar zu erwerben. Für die Präsentation und Diskussionsbeiträge können Bonuspunkte erworben werden. In der Seminararbeit sind maximal 75 Punkte zu erreichen, durch die mündliche Präsentation und Diskussion zusätzlich insgesamt maximal 25 Bonuspunkte. Die Modulnote ergibt sich aus einer Punkteskala, die von 0 bis 100 Punkte reicht. Die Bonuspunkte werden nur angerechnet, wenn die Seminararbeit auch ohne die Bonuspunkte bestanden wäre.

Credit-Points

Bei erfolgreicher Teilnahme an dem Kurs werden 5 Credit-Punkte vergeben.

Seminarthemen

Bitte achten Sie darauf, in Ihrer Bewerbungs-Email Ihren Erst-, Zweit- und Drittwunsch für ein Seminarthema zu nennen (mit Angabe der Nummern 1. bis 10.). Nach Anmeldungsschluss wird Ihnen ein Thema zugeteilt. Es kann dabei nicht garantiert werden, dass jeder das zuerst gewünschte Thema erhält. Wenn mehrere Studierende dasselbe Thema als Erstwunsch angeben, wird dieses an eine der Personen nach dem Zufallsprinzip zugeteilt. Dann erfolgt die Zuteilung des Zweitwunsches, usw. Jedes Thema kann maximal zweimal verteilt werden. Jede/r Studierende schreibt eine eigene Seminararbeit. Das zugeteilte Thema wird Ihnen während der Einführungsveranstaltung mitgeteilt. Sollte ein Thementausch erwünscht sein, so muss sie/er dazu eine/n andere/n Tauschpartner*in finden. Ein Thementausch kann während der Einführungsveranstaltung stattfinden.

Die Studierenden sollen in ihrer Seminararbeit relevante Theoriemodelle aus der Data Science, die in den ihnen zugeteilten Seminarthemen (siehe list of topics and research papers) angewendet werden, erläutern. Theorien aus der Data Science finden sich u.a. in den folgenden Textbüchern:

- a) **James, Witten, Hastie, Tibshirani, Taylor (2023), An Introduction to Statistical Learning – With Applications in Python**, downloadbar unter https://hastie.su.domains/ISLP/ISLP_website.pdf
- b) **Jurasky, Martin (2021), Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**, 3. Auflage, downloadbar unter <https://web.stanford.edu/~jurafsky/slp3>

Die genannte Literatur zu den einzelnen Seminarthemen gilt als Startpunkt für die eigene Literatursuche. D.h., für das Bestehen des Seminars ist es erforderlich, dass Sie sich zum einen mit dem Ihnen zugeteilten Forschungspapier auseinandersetzen und wichtige Elemente daraus in Ihrer Seminararbeit beschreiben, zum anderen werden Sie darauf aufbauend weitere Literatur hinzuziehen. Wir werden über weitere Literatur für Ihre Seminararbeit auch bei unseren Treffen sprechen.

List of topics and research papers:

1. Environmental Economics

- Ma, N.; Shum, W.; Han, T.; Lai, F. (2021). Can Machine Learning be Applied to Carbon Emissions Analysis: An Application to the CO2 Emissions Analysis Using Gaussian Process Regression, *Frontiers in Energy Research*, 9, 756311.

2. Development Economics

- Baker, S.; Bloom, N.; Davis, S. (2016). Measuring Economic Policy Uncertainty, *Quarterly Journal of Economics*, 131, 4.

3. Development Economics

- Fetzner, T. (2019). Can Workfare Programs Moderate Conflict? Evidence from India, *Journal of the European Economic Association*, 18(6), 3337–3375.

4. Entrepreneurship

- Coad, A.; Srhoj, S. (2020). Catching Gazelles with a Lasso: Big data techniques for the prediction of high-growth firms, *Small Business Economics*, 55, 541-565.

5. Gender Economics

- Davis, S.; Worsnop, C.; Hand, E. (2022). Gender bias recognition in political news articles, *Machine Learning with Applications*, 8, 100304.

6. Regional Economics

- Mikko, M.; Stein, O.; Jaakko, S. (2022). Machine learning and the identification of Smart Specialisation thematic networks in Arctic Scandinavia, *Regional Studies*, 56, 9, 1429-1441.

7. International Economics

- Lu, Y.; Shao, X.; Tao, Z. (2018). Exposure to Chinese imports and media slant: Evidence from 147 U.S. local newspapers over 1998-2012, *Journal of International Economics*, 114, 316-330.

8. Labour Economics

- Kässä, O.; Lehdonvirta, V. (2018). Online labour index: Measuring the online gig economy for policy and research, *Technological Forecasting and Social Change*, 137, 241-248.

9. Labour Economics

- Ros, R.; Erp, M.; Rijpma, A.; Zijdeman R. (2020). Mining Wages in Nineteenth Century Job Advertisements. The Application of Language Resources and Language Technology to study Economic and Social Inequality, in Proceedings of the Workshop about Language Resources for the SSH Cloud, pp. 27-32.

10. Monetary Economics

- Hansen, S.; McMahon M.; Tong, M. (2019). The long-run information effect of central bank communication, *Journal of Monetary Economics*, 108, 185-202.