Studiengang Informationswissenschaft (Bachelor of Science)

Wahlpflichtkatalog

Themenbereich: Wirtschaftsinformation

Modul-bezeichnung	Methoden-Workshop: Business Intelligence 1.0 und 2.0 (Methods-Workshop: Business Intelligence 1.0 and 2.0)
Belegnummer	7507
Studiengang / Verwendbarkeit	Bachelorstudiengang Informationswissenschaft
Modulverantwort-liche(r)	Prof. Dr. Bernd Jörs
Dozent(in)	Prof. Dr. Bernd Jörs
Dauer	1 Semester
Credits	5 CP
Prüfungsart	Prüfungsleistung: Klausur oder mündliche Prüfung
Sprache	deutsch
Inhalt	Die Arbeitswelt der Bachelorabsolventen wird derartige Qualifikationen gerade im Bereich des „Information Science and Engineering“ und der „Business Intelligence“ vermehrt abverlangen. • Nicht zuletzt das „Engineering“, z.B. die Auseinandersetzung mit Fragen des Design und der Ausformung von elektronischen Märkten bzw. Plattformen im Sinne des Market Engineering, ein Bestandteil des „Business Information Engineering“ erfordert ein Verständnis des „Economist as Engineer“, d.h. jemand, der befähigt ist, elektronische Märkte „mit ingenieurwissenschaftlichen Ansätzen und Verfahren in Verbindung zu bringen“ und dabei explizit auch das (rationale und irrationale) Information Behavior und Decision Behavior der Marktteilnehmer berücksichtigt, wie Prof. Veron Smith, Nobelpreisträger Wirtschaft 2002) forderte. • Modernes Web-Controlling im Sinne der immer wichtiger werdenden „Web Analytics“ – also zeitgemäße Nutzerverhaltensforschung, Click-stream-analysis, Tracking ist ohne Kenntnis von quantitativen Datenauswertungsmethoden und –tools nicht möglich, um z.B. die stichprobenartigen Nutzerprofile auf ihre „Signifikanz“ oder „Konfidenz“ zu überprüfen. Gleiches gilt auch für die „user experience“ (UX) und usability-Forschung im Rahmen der interdisziplinären Fachrichtung der „Informations-Architektur“. Wie wird hier methodisch einwandfrei „experimentiert“ und „getestet“? • Experimentieren, Simulationen durchführen, Testverfahren anwenden oder empirische Feldforschung betreiben – all dies muss auf ein methodisch sauberes und nachvollziehbares Fundament gestellt werden. So verlangt z.B. das dem market engineering zugeordnete Planen und Kontrollieren von so genannten „Empfehlungssystemen“ (recommendation systems), bestes Beispiel die fortlaufenden, personalisierten Empfehlung auf der Website von Amazon, das stichprobenartige Testen von derartigen Kaufempfehlungen auf Nutzerrelevanz und –akzeptanz, um die Such- und Entscheidungsprozesse für den Nutzer zu optimieren.

- Im Rahmen des Online Marketing, hier zum Beispiel bei der Anwendung von multivariaten Verfahren der Landingpage-Optimierung, der strategischen und operativen Unternehmensplanung, der Marktforschung, der Kosten- und Erlös- bzw. Budgetschätzung oder der Investitions- und Finanzrechnung bzw. Kapitalmarkt-Risikoanalyse sind methodisch einschlägige Datenanalysen und -prognosen und deren Qualität das non-plus ultra. Wie erhält man qualitativ gute und methodisch akzeptierte Vorhersagen?
- Die methodisch professionelle, mathematisch-statistisch akzeptierte Aufbereitung und „zielführende Gestaltung bzw. Interpretation“ der Ergebnisse sind auch Gegenstand des gesamten Anwendungsfeldes der Datenanalytik und der damit immer stärkeren Datenvermarktungswirtschaft, z.B. im Bereich der Sportdatenerhebungen, Medienanalysen, Geo-Daten, Facebook- oder Google-Datenanalyse etc.
- Wissenschaftliche Messmethodenfragen aus Sicht der social media network analysis und die einführende Auseinandersetzung mit Fragen der (statistischen) Datenerhebung und –auswertung im Rahmen der Wirtschaftlichkeits- und Nutzerverhaltensanalysen runden die Thematik ab.

Gliederung:

A. Aus (Trainings-)Daten „lernen“: Regeln (Rules) und Muster (Pattern)

- 1. Informationsökonomische Grundlagen
  - 1.1. Informationstheorie von C. Shannon
  - 1.2. Entropie – der mittlere Informationsgehalt
  - 1.3. Informationsgewinn (Information Gain)
  - 1.4. Klassifikation per Entscheidungsbaum (Decision Tree Learning)
  - 1.5. Anwendungsfälle des ID3-Algorithmus für Decision Trees
  - 1.6. Regressionsbäume – ein besondere Variante der Decision Trees
  - 1.7. Ergebnisüberprüfung mithilfe der Mean Squared Error (MQF)-Methode
- 2. Maschinelles Lernen
  - 2.1. Ziele, Begriffe und Anwendungsfelder des Maschinellen Lernens
  - 2.2. Regeln lernen und Muster erkennen: Supervised / unsupervised learning
  - 2.3. Beispielsalgorithmen: Find-S-Algorithm, Version Space, Find-G-Set-Algorithm
  - 2.4. Optimierungsansätze: Post-Pruning (Candidate Eliminate-Algorithm)
  - 2.5. Regel-Algorithmus: Separate-and-Conquer-Rule (Top-Down-Hill-Climbing)
  - 2.6. Precision & Accuracy
  - 2.7. Bottom-Up-Hill-Climbing (Special-to-General)
  - 2.8. Evaluation der Regeln (Learning outcomes): Sign Test, Covering-Algorithm

B. Daten-Ähnlichkeiten und Daten-Klassen (Cluster)

- 3. Ähnlichkeitsmessung (similarity measurement) zur Mustererkennung
  - 3.1. Binäre Ähnlichkeitsmaße (Jaccard-, M-Koeffizient etc.)
  - 3.2. Nominale Ähnlichkeitsmaße (Distanzmaße, L-Norm, Eulidische Distanz, Q-Korrelationskoeffizient etc.)
  - 3.3. Cluster-Analytik (agglomerative, divisive Clusteralgorithmen)
  - 3.4. Maschinelles Lernen und Cluster-Analytik (mit Inverse Distance Weighting, Value Difference, Metric VDM, RelieF-Algorithm zur Attributgewichtung, RISE, Initiale Accuracy)
  - 3.5. Chi-Quadrat und Korrespondenzanalyse (+ Maschinelles Lernen mit supervised/unsupervised discretization)
  - 3.6. Assoziationsanalyse (Warenkorbanalyse)
  - 3.7. Informationswissenschaftliche Ähnlichkeitsmessung (Vektorraummodell und WDF-IDF-Modell)
  - 3.8. Anwendungsfälle der Ähnlichkeitsmessung (Empfehlungssysteme/Recommerdersystems; content-based-filtering, collaborative filtering)

C. Daten-Verbindungen im Social Web

- 4. Vernetzungsmessung (social media network analysis)
  - 4.1. Interconnectedness
  - 4.2. Zentralitätsmaße (Degree Centrality, Closeness Centrality, Betweenness Centrality, Influence measures)
  - 4.3. Social Media Page Rank

D. Daten-Prognose

- 5. Einfache Prognoseverfahren
  - 5.1. Zeitreihenanalyse und Trendberechnung
  - 5.2. Regressionsanalyse (bi- und multivariate Analyse)
  - 5.3. Neuere (qualitative) Prognostik (Conjoint Measurement, Cross-Impact-Analysis, Take-the-Best-Heuristik)
  - 5.4. Webbasierte Analyse- und Prognoseverfahren (Prediction Markets)
  - 5.5. Anwendungsfälle aus dem Finanzmanagement (Corporate Finance)
  - 5.5. Kritische Prüfung

Angestrebte Lernergebnisse (Learning Outcome)	Das pure Anwenden von professioneller Anwendersoftware für die Erhebung, Aufbereitung und Auswertung von (Unternehmens-)Daten aus dem Hause IBM, Oracle oder SAP zur Datenanalyse und –aufbereitung erleichtert heutzutage im Zeitalter der Massendaten und „Big Data"-Diskussion die Bearbeitung, Untersuchungen und Auswertungen von data and information overload in Unternehmen und Institutionen. Schon Microsoft bietet mit ihrem MS SQL Server integrierte Data-Mining-Komponenten für alle Auswertungen der verschiedensten Daten an. „Als Benutzerschnittstelle für die Vorbereitung der Daten und Nutzung der Modelle gibt es für MS Excel die Data-Mining-Adds-Ins." Das ist alles sehr schön, professionell und sehr hilfreich. **Aber...** Die effiziente Befassung mit Daten und deren Ergebnisinterpretationen sowie eine kritisch-rationale Sicht auf die Datenergebnisse und ein Schutz vor einer kritiklosen Übernahme der per Software ermittelten Datenauswertungen erfordert eines ganz bestimmt: **METHODEN-KENNTNISSE.** Um die bestehenden Fähigkeiten der Studierenden zu erweitern und zusätzliche, praxistaugliche Alleinstellungsmerkmale zu vermitteln, soll in diesem Fachmodul ein **methodisch-operatives Rüstzeug** im Umgang mit Business Intelligence-, insb. Data- und Web Mining-Verfahrenstechniken bzw. Maschinellen Lernverfahrensansätzen mit auf den Weg gegeben werden. AbsolventInnen des Bachelorstudiengangs sollen später schnell, kostengünstig und zielführend für verschiedene Kunden, Nutzer und Entscheider Informations(vermarktungs)dienstleistungen auf hohem qualitativem und wissenschaftlichem Niveau vollbringen. Auf einem der wichtigsten beruflichen Arbeitsfelder der Zukunft, der Aufbereitung von **strukturierten** und **unstrukturierten** (Massen-)**Daten**, nicht zuletzt durch die aufgekommene „Big Data"-Diskussion angestoßen, sind zur Erlangung von arbeitsmarktrelevanten, wettbewerbsfähigen Qualifikationsalleinstellungsmerkmalen u.a. gute **methodische skills** zur Analyse derartiger strukturierter und unstrukturierter Datenmengen dringend notwendig. Dazu muss man u.a. auf die in der „scientific und practice community" bekannten und akzeptierten quantitativ-qualitativen, heuristisch-statistischen Verfahren zurückgreifen. Aber dies nicht kritiklos und „blind". Das moderne Management benötigt Mitarbeiter, die fundierte (empirische) Analyse-, Klassifikations- und Prognosemethoden kennen und beherrschen, aber auch deren Aussagekraft und Grenzen bei der Datenerhebung, -aufbereitung, -analyse und –aufbereitung richtig einschätzen können; gerade im Zeitalter der (webbasierten) Massendatenproduktion („Big Data") ist hier ein kritisch-wacher Sachverstand notwendig, denn die Ankündigungen sind beeindruckend: • **"Data is the new oil"** (Gerd Leonhard, The Media Futurist) **Data will become a key currency, as it is a virtually limitless, non-rival, and exponentially growing good. What will Generation AO (always-on) share with whom, when, where, and how? Data is exploding all around us: every 'like,' check-in, tweet, click, and play is being logged and mined. Many data-centric companies such as Google are already paying us for our data by providing more or less free services.** • **„The sexiest job in the next 10 years will be statisticians. People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s. If „sexy" means having rare qualities that are much in demand, data scientists are already there"** (Prof. Dr. Hal Varian, Chief Economist Google Inc.) • **Data Scientist: The Sexiest Job of the 21st Century by Thomas H. Davenport and D.J. Patil Data Scientist: The Sexiest Job of the 21st Century** (Thomas H. Davenport and D.J. Patil, Harvard Business Review 10/2012) • **Are you ready for the era of 'big data'? : Radical customization, constant experimentation, and novel business models will be new hallmarks of competition as companies capture and analyze huge volumes of data.** (McKinsey&Company 2012)

Carolin Kaiser stellt in ihrem Buch „Business Intelligence 2.0“ die richtigen Ausgangsfragen:

- Wie kann wertvolles Wissen aus dem (Web 1.0, der Verf.) und Web 2.0 gewonnen werden? (Mining-Services)
- Wie kann dieses Wissen über die Zeit hinweg überwacht werden? (Monitoring-Services)
- Wie kann frühzeitig von kritischen Situationen gewarnt werden? (Frühwarn-Services)
- Wie können Entscheidungen zur Meinungsbeeinflussung unterstützt werden ? (Entscheidungsunterstützung-Services)

Was sind das für **Analyse-, Klassifikations- und Vorhersagemethoden**, was können sie und was können sie nicht?

Warum wird über eine vereinfachte Darstellung nicht die eigentliche (begrenzte) Substanz dieser oft sehr mathematisch formelhaft komplex dargestellten Methoden offen gelegt, wie in diesem Fachmodul vorgesehen?

Will man durch formelhafte Berechnungskomplexität und komplizierte Herleitung wissenschaftlich beeindrucken, nach der Devise: Je schwieriger und schwerverständlich, desto besser die Analyse-, Klassifikations- und Prognosequalität? Baue ich hier eine eigene (fiktive, realitätsferne) Wissenschaftswelt auf, die lediglich dem armseligen „Beeindrucken“ gilt, die häufig dogmatisch und autoritär erscheint, statt dem eigentlichen Ziel, die ökonomische und soziale Realität zu erklären und zu prognostizieren?

Warum fällt es so schwer, sich neuen Erkenntnissen und Verfahrenstechniken der qualitativ-intuitiven Prognostik oder der **webbasierten Datenerhebungs- und -analysetechniken** für die Analyse-, Forschungs- und Prognosearbeit zu öffnen, die nachweislich bessere Ergebnis- und Vorhersagequalitäten besitzen, wie Auswertungen bei Google Analytics oder elektronischen Plattformen wie „prediction markets“ belegen ?

Im Fokus der Lehrveranstaltung steht das Qualifikationsziel der anwendungsorientierten Vermittlung von Verfahrenstechniken des empirisch-experimentellen Data- und Web-Mining, insbesondere mit Bezug auf die Grundlagen Maschinellen Lernens (als Bestandteil des Knowledge Discovery in Databases KDD).

Ausgangspunkt sind die methodischen Anaylsetechniken des **Data-Mining**, das versucht – wie in Wikipedia allgemein formuliert – „aus einem Datenberg etwas Wertvolles (zu) extrahieren“. Methodenbasis für eine systematische Auswertung der Daten, die häuifg wertvolles implizites Wissen enthalten, ist die Anwendung bestimmter, anerkannter deskriptiver und induktiver statistischer Analyseverfahren „mit dem Ziel, neue Muster zu erkennen.“ Text- und Web-Mining nutzen diese methodischen Grundlagen des Data Mining, um solche Muster (pattern) aus eher unstrukturierten Daten herauszufiltern.

Wie lassen sich aus Vergangenheitsdaten (Trainingsdaten) **Regelhaftigkeiten, Muster, Zusammenhangs- und Abhängigkeitsbeziehungen, Prognosepotenziale, Ähnlichkeiten, Klassifikationen (Cluster, Assoziationen) oder Netzwerkverbindungen** herleiten und anhand von Testdaten sowie durch überwachtes oder nicht-überwachtes maschinelles Lernen überprüfen?

Wie wird dies methodisch realisiert ? Kann man damit gute Vorhersagen machen?

	Es bedarf also dreier grundsätzlicher Qualifikationsziele: 1. Befähigung zum Umgang mit quantitativ-qualitativen, heuristisch-statistischen Verfahren des Data- und Web Mining als Methodentools der Web Science 2. Anwendungsbefähigung und Verständnisschaffung für die Nutzung einschlägiger Anwenderstandardsoftware (z.B. die weltweit mit am häufigsten zur Anwendung kommende IBM SPSS Modeller Software, die an der Hochschule als Testsoftware mit nahezu allen Funktionalitäten für Studenten des Studiengangs zur Verfügung steht) 3. Kritisch-rationale Einschätzung der Möglichkeiten und Grenzen der Anwendung und Aussagekraft herkömmlicher und neuer Analyse-, Forschungs- und Prognosemethoden. Die Vermittlung mathematisch/heuristischer-statistischer, insb. „multivariater Verfahren“, löst oftmals ein „ungutes“ Gefühl aus, deshalb werden stellen sich für den Dozenten besondere Herausforderungen. Dies erfolgt in Form einer „**Anti-Hegel**“-Lehrveranstaltung: *„Er hat dazu geführt, dass es in Universitäten – in vielen Universitäten, natürlich nicht in allen – eine Tradition gibt, Dinge hegelianisch auszudrücken, und dass die Leute, die das gelernt haben, es nicht nur als ihr Recht ansehen, so zu sprechen, sondern geradezu als ihre Pflicht. Aber diese sprachliche Einstellung, die Dinge schwierig und damit eindrucksvoll auszudrücken, die macht die deutschen Intellektuellen unverantwortlich. . . Die intellektuelle Verantwortlichkeit besteht darin, eine Sache so deutlich hinzustellen, dass man dem Betreffenden, wenn er etwas Falsches oder Unklares oder Zweideutiges sagt, nachweisen kann, dass es so ist“ Es gibt eine Art Rezept für diese Dinge:. . . Man sage Dinge, die großartig klingen, aber keinen Inhalt haben, und gebe dann Rosinen hinein – die Rosinen sind Trivialitäten. Und der Leser fühlt sich gebauchpinselt, denn er sagt, das ist ja ein ungeheuer schweres Buch!* (Sir Karl Popper 1990) Die Lehrveranstaltung soll daran gemessen werden, ob sie den kritisch-rationalen Anmerkungen von Karl Popper Folge geleistet haben.
Niveaustufe / Level	Fortgeschrittenes Niveau (advanced level course)
Lehrform / SWS	Seminar (4 SWS)
Arbeitsaufwand / Workload	128 Stunden
Units (Einheiten)	
Notwendige Voraussetzungen	
Empfohlene Voraussetzungen	Interesse an einer methodisch-wissenschaftlichen Qualifikation für Aufgaben im Business Intelligence-, Online-Marketing-, Wirtschafts- und Finanz-, Marktforschungs- oder Wissenschaftsbereich **Da die Lehrveranstaltung als (geblockter) Methodenworkshop angeboten werden soll und die Teilnehmer schon während der Veranstaltung die Anwendung der Methoden üben sollen, wird die Bereitschaft zur aktiven und ernsthaften Teilnahme eine elementare Voraussetzung sein. Interessenten, die andere für sich arbeiten und rechnen lassen wollen, in der Lehrveranstaltung lieber online googeln, sollten diese Lehrveranstaltung nicht belegen.**
Häufigkeit des Angebots	
Anerkannte Module	Siehe § 19 ABPO

Medienformen	
Literatur	• Oestreich, M.; Romberg, O.: Keine Panik vor Statistik! Erfolg und Spaß im Horrorfach nichttechnischer Studiengänge. 3.Aufl., 2010 • Monka, M.; Schöneck, Nadine M.; Voss, W.: Statistik am PC. München 2008 • Caputo, A.; Fahrmeier, L.; Künstler, R.; Pigeot, I.; Tutz, G.: Arbeitsbuch Statistik. 5.Aufl., Berlin 2008 • Bamberg, G.; Baur, F.; Krapp, M.: Statistik-Arbeitsbuch. Übungsaufgaben, Fallstudien, Lösungen. 8.Aufl., München 2007 • Wewel, Max.: Statistik im Bachelor-Studium der BWL und VWL. München 2011 • Krämer, W.: So lügt man in der Statistik. 2011 • Alpaydin, Ethem: Maschinelles Lernen, Oldenbourg-Verlag, 2008 • Witten, Ian H.; Frank, Eibe; Hall, Mark A.: Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 3.Aufl., 2011 • Ferber, Reginald: Information Retrieval. Suchmodelle und Data Mining für Textsammlungen und das Web, Dpunkt Verlag, 2003 • Russell, Matthew: Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and other Social Media, O'Reilly Media, 2011 • Kemper, Hans-Georg; Baars, Henning; Mehanna, Walid: Business Intelligence – Grundlagen und praktische Anwendungen, Vieweg+Teubner, 3.Aufl., 2010 • Skulschus, Marco; Tittel, Jan; Wiederstein, Marcus: MS SQL Server – Data Mining, Analyse und multivariate Verfahren; Comelio Medien, 2013 Zusätzliche Unterlagen, Übungsaufgaben und Materialen.

Stand: 12.09.2014, 14:23:54