Oregon State University
Ecampus

Course Name: Data Analytics and Applied Machine Learning
Course Number: CS 419 ST
Credits: 4
Instructor name: Rick Hangartner
Instructor email: rick.hangartner@oregonstate.edu
Teaching Assistant name and contact info: Jing Wang
(wangji9@oregonstate.edu), Ajay Krishna (krishnaj@oregonstate.edu)

Course Description
This course offers an introduction to data analytic concepts and machine learning techniques used in business and research. Prerequisite: CS 225 or equivalent.

Communication
Please post all course-related questions in the course Piazza Discussion Forums so that the whole class may benefit from the conversation. Please contact me privately for matters of a personal nature. I or a course TA will endeavor to reply to course-related questions within 24 hours. I will strive to return your assignments and grades for course activities to you within five days of the due date.

Course Credits
This course combines approximately 120 hours (about 12 hours per week) of instruction, online activities and assignments for 4 credits.

Technical Assistance
If you experience any errors or problems while in your online course, contact 24-7 Canvas Support through the Help link within Canvas. If you experience computer difficulties, need help downloading a browser or plug-in, or need assistance logging into a course, contact the IS Service Desk for assistance. You can call (541) 737-8787 or visit the [IS Service Desk](#) online.

Learning Resources
This course has two required, widely available texts:
- *Data Science for Business*, Foster Provost and Tom Fawcett, O'Reilly, 2013. ISBN 9781449361327. See also:
  https://www.oreilly.com/catalog/errata.csp?isbn=9781449361327
- *Practical Statistics for Data Scientists*, Peter Bruce and Andrew Bruce, 2017. ISBN 9781491952962. See also:
  https://www.oreilly.com/catalog/errata.csp?isbn=9781491952962
- *Statistical Methods 3rd Ed.*, Rudolf J. Fruend, Donna Mohr, and William J. Wilson, 2010. (Reference) http://OSU.eblib.com/patron/FullRecord.aspx?p=802390

We use two IBM software systems:
- IBM SPSS Modeler: Available online. On OSU-EECS Citrix Workspaces with your ONID ID
- IBM Watson Studio Cloud: Available online. You'll sign up for an IBM Cloud account using your OSU email address

You can install IBM SPSS Modeler and IBM Watson Studio Local version on your own machine, if you prefer, but EECS IT cannot support any installation on your own machine.

Note to prospective students: Please check with the OSU Beaver Store for up-to-date information for the term you enroll (OSU Beaver Store website or 800-595-0357). If you purchase course materials from other sources, be very careful to obtain the correct ISBN.

Measurable Student Learning Outcomes

At the completion of the course, students will be able to:
- **Identify** problems addressable by data-driven solutions
- **Develop** datasets that can be used for building machine learning-enabled technologies
- **Describe** the relationship between descriptive/explanatory modeling in statistics and predictive modeling in machine learning.
- **Build** data models for predictive analytics using machine learning tools.
- **Determine** the mix of machine learning techniques applicable for a problem.
- **Evaluate** the performance of machine learning models and data analytic techniques.

Evaluation of Student Performance

This is a computing laboratory course with significant weekly assignments and reports, along with class discussion.
- Discussions - 400 points
- Assignments - 450 points
- Terminology Quiz (Week 2) - 20 points
- Resource Monitoring Assignments - 100 points
- Total - 970 points

Letter Grade

Grade	Percent Range	Points
A	94-100	911-970
A-	90-93	873-910
B+	87-89	843-872
B	84-86	814-842
B-	80-83	776-813
C+	77-79	746-775
C	74-76	717-745
C-	70-83	679-716
D+	67-69	649-678

D	64-66	620-648
D-	60-63	582-619
F	59 or less	0-581

Course Content

Week	Topic	Reading Assignments	Learning Activities
1	Computational Tools and ML Platforms	**Disc 2:** Walch - "The future of data science and AI points to automatic tools"	**Exp 1:** Technology Overview **Exp 2:** Data Analytics Software **Exp 3:** Data Science Platforms **Disc 1:** Introductions all around **Disc 2:** What is the future of automated AI? **Assign 1:** Introducing IBM SPSS Modeler **Assign 2:** Introducing IBM Watson Studio Cloud **Assign 3:** Monitoring Cloud Usage
2	Data Science and Data Analytic Thinking	**Exp 2:** Provost and Foster, Ch. 1- "Introduction: Data Analytic Thinking" **Exp 3:** Provost and Foster, Ch. 2 - "Business Problems and Data Analytic Solutions" **Disc 1:** Stirrup - "What's wrong with CRISP-DM, and is there an alternative?" **Disc 2:** Bayern - "How to fail as a data scientist: 3 common mistakes"	**Exp 1:** Introduction to Data Science **Exp 2:** Data Science Terminology **Exp 3:** Data Mining **Disc 1:** Is CRISP-DM obsolete? **Disc 2:** What do you hope to learn in this class? **Assign 1:** Terminology Worksheet **Assign 2:** Monitoring Cloud Usage
3	Working with datasets	**Exp 1:** Breiman - "Statistical Modeling: Two Cultures" **Exp 2:** Zuboff C-SPAN interview - "The Rise of Surveillance Capitalism" **Exp 3:** Provost and Foster, Ch. 2, pp. 30-31 - "Data Preparation"	**Exp 1:** About data **Exp 2:** A closer look at behavioral data **Exp 3:** Data preparation **Disc 1:** Data science for fun and profit? **Disc 2:** Data science for bad or good? **Assign 1:** Handling Missing Feature Values in a Dataset

		Provost and Foster, Ch. 5, pp. 113-115 - "Holdout Data and Fitting Graphs" Provost and Foster, Ch. 5, pp. 126-129 - "From Holdout Evaluation to Cross-Validation" Bruce and Bruce, Ch. 4, pp. 138-139 - "Cross-Validation" **Disc 1:** Fogg - "Anthony Goldbloom gives you the secret to winning Kaggle competitions" **Disc 2:** Elgan - "Uh-oh: Silicon Valley is building a Chinese-style social credit system"	**Assign 2:** 10-Fold Partitioning of a dataset **Assign 3:** Monitoring Cloud Usage
4	Basic Statistical Techniques for Machine Learning	**Exp 1:** Bruce and Bruce, Ch. 1 - "Exploratory Data Analysis" **Exp 3:** Bruce and Bruce, Ch. 2 - "Data and Sampling Distributions" **Disc 1:** Provost and Fawcett, Ch. 3, pp 44-47 - "Models, Induction, and Prediction" Bruce and Bruce, Ch. 4, pp. 133-134 - "Prediction vs. Explanation" Bruce and Bruce, Ch. 4, pp. 142-145 - "Prediction Using Regression" **Disc 2:** Greengard - "An Inability to Reproduce"	**Exp 1:** Exploring Data **Exp 2:** The Big Idea **Exp 3:** About Samples **Disc 1:** What is Prediction? **Disc 2:** The risk as two cultures converge? **Assign 1:** Practical Statistics and Statistical Tools for Numeric Data **Assign 2:** Statistical Tools for Categorical Data **Assign 3:** Monitoring Cloud Usage
5	Supervised Learning: Algorithmic Models for Prediction	**Exp 1:** Provost and Foster, Ch. 3 - "Introduction to Predictive Modeling: From Correlation to Supervised Segmentation"	**Exp 1:** Decision Trees **Exp 2:** Decision Trees as Predictive Models **Exp 3:** Ensembles of Decision Trees **Disc 1:** What does ML learn? **Disc 2:** What's the Problem?

		Disc 1: Marcus and Davis - "How to Build AI We Can Trust" **Disc 2:** The Economist - "Trouble at the Lab"	**Assign 1:** Decision Tree Models **Assign 2:** Ensemble Models **Assign 3:** Monitoring Cloud Usage
6	Supervised Learning: Data Models for Explanation	**Exp 1:** Provost and Foster, Ch. 4 - "Fitting a Model to Data" **Disc 1:** Campbell - "Why Advertising Doesn't Work Very Well, People Profiled by Demographics Actually Agree on Very Little" **Disc 2:** Sullivan - "There could be a simple solution to Facebook's political ad woes"	**Exp 1:** Regression Models for Numeric Data **Exp 2:** Discriminant Models for Categorical Data **Exp 3:** Regression Models for Categorical (Nominal) Data **Disc 1:** Identity is Not Enough? **Disc 2:** This Would Work How? **Assign 1:** Logistic Regression Models **Assign 2:** Comparing Logistic Regression Models to Tree Models **Assign 3:** Monitoring Cloud Usage
7	Exploring the Limits of Models	**Exp 1:** Provost and Foster, Ch. 5 - "Overfitting and Its Avoidance" **Disc 1:** Panch, Mattie, and Celi - "The 'inconvenient truth' about AI in healthcare" **Disc 2:** Zhou and Sun - "A Case Against Mission-Critical Applications of Machine Learning"	**Exp 1:** Introduction to Model Fitting **Exp 2:** Formal Definition of Model Fit **Exp 3:** Practical Details of Fitting Models **Disc 1:** When is the Data Suitable? **Disc 2:** What's the Risk? **Assign 1:** Assessing Model Fit **Assign 2:** Understanding Model Validation **Assign 3:** Monitoring Cloud Usage
8	Unsupervised Learning: Models for Discovery	**Exp 1:** Provost and Foster, Ch. 6 - "Similarity, Neighbors, and Clusters" **Disc 1:** Davies - "The Washington Post is preparing for post-cookie ad targeting" **Disc 2:** Price, Gerke, and Cohen - "Potential Liability for Physicians	**Exp 1:** Distance as the Core Principle of Clustering **Exp 2:** Clustering for Classification **Exp 3:** Clustering for Pattern Discovery **Disc 1:** Shaping the News? **Disc 2:** What if Bad ML Mattered? **Assign 1:** Introduction to Unsupervised Learning

5

		Using Artificial Intelligence"	**Assign 2:** Investigating Unsupervised Learning **Assign 3:** Monitoring Cloud Usage
9	From Predictive Analytics to Decision Analytics	**Exp 1:** Provost and Foster, Ch. 7 - "Decision Analytic Thinking I: What is a Good Model" **Disc 1:** Cahan - "Putting the data before the algorithm in big data addressing personalized healthcare" **Disc 2:** Denning and Lewis - "Intelligence May Not be Computable"	**Exp 1:** Classifier Performance Evaluation **Exp 2:** Dataset Priors and Classifier Metrics **Exp 3:** Performance Evaluation as Expected Value **Disc 1:** Good Algorithms or Good Data? **Disc 2:** Do Data and Algorithms make ML Intelligence? **Assign 1:** Classifier Evaluation Metrics **Assign 2:** Effects of Dataset Imbalance on Classifier Metrics **Assign 3:** Monitoring Cloud Usage
10	Using Decision Analytics	**Exp 1:** Provost and Foster, Ch. 8 - "Visualizing Model Performance" **Disc 1:** Health Catalyst, Inc - "Machine Learning and Feature Selection for Population Health" **Disc 2:** Lecher - "A health care algorithm affecting millions is biased against black patients"	**Exp 1:** Classifier Scores **Exp 2:** Classifier Decision Threshold Optimization **Exp 3:** A Final Word **Disc 1:** What Makes Useful Predictive Applications in Practice? **Disc 2:** Algorithm or Application? **Assign 1:** Classifier Performance Visualization **Assign 2:** Classifier Performance Optimization **Assign 3:** Monitoring Cloud Usage
Finals	N/A	N/A	N/A

Course Policies

Discussion Participation
The collaborative, graded class Discussions are a significant part of this course. Students are expected to participate in all graded discussions. While there is great flexibility in online courses, this is not a self-paced course. You will need to participate in discussions on at least two different days each week, with your first post due no later than Wednesday evening, and your second (and any optional additional posts) due by the end of each week. Employers consistently say that they value most data scientists and machine learning specialists who not only have technical expertise but who can communicate about the value

of the analyses and solutions they create to non-technical team members. The Discussion component responds to this opportunity.

Late Work Policy
This is a fast-paced course with weekly discussions and assignments. Late work will only be accepted only in emergency cases (usually only for a death in the family, major illness or injury, or birth of your child).

Proctored Exams and Makeup Exams
There are no exams in this course.

Incompletes
Incomplete (I) grades will be granted only in emergency cases (usually only for a death in the family, major illness or injury, or birth of your child), and if a student has turned in work worth 80% of the points possible. If you are having any difficulty that might prevent you from completing the coursework, please don't wait until the end of the term; let me know right away.

Guidelines for a Productive and Effective Online Classroom
Students are expected to conduct themselves in the course (e.g., on discussion boards, email) in compliance with the university's regulations regarding civility. Civility is an essential ingredient for academic discourse. All communications for this course should be conducted constructively, civilly, and respectfully. Differences in beliefs, opinions, and approaches are to be expected. In all you say and do for this course, be professional. Please bring any communications you believe to be in violation of this class policy to the attention of your instructor.

Active interaction with peers and your instructor is essential to success in this online course, paying particular attention to the following:
- Unless indicated otherwise, please complete the readings and view other instructional materials for each week before participating in the discussion board.
- Read your posts carefully before submitting them.
- Be respectful of others and their opinions, valuing diversity in backgrounds, abilities, and experiences.
- Challenging the ideas held by others is an integral aspect of critical thinking and the academic process. Please word your responses carefully, and recognize that others are expected to challenge your ideas. A positive atmosphere of healthy debate is encouraged.

Reach Out for Success
University students encounter setbacks from time to time. If you encounter difficulties and need assistance, it's important to reach out. Consider discussing the situation with an instructor or academic advisor. Learn about resources that assist with wellness and

7

[academic success](#). If you are in immediate crisis, please contact the Crisis Text Line by texting OREGON to 741-741 or call the National Suicide Prevention Lifeline at 1-800-273-TALK (8255).

Statement Regarding Students with Disabilities
Accommodations for students with disabilities are determined and approved by Disability Access Services (DAS). If you, as a student, believe you are eligible for accommodations but have not obtained approval, please contact DAS immediately at 541-737-4098 or at http://ds.oregonstate.edu. DAS notifies students and faculty members of approved academic accommodations and coordinates implementation of those accommodations. While not required, students and faculty members are encouraged to discuss details of the implementation of individual accommodations.

Accessibility of Course Materials
All materials used in this course are accessible. If you require accommodations please contact Disability Access Services (DAS).

Additionally, Canvas, the learning management system through which this course is offered, provides a vendor statement certifying how the platform is accessible to students with disabilities.

Expectations for Student Conduct
Student conduct is governed by the university's policies, as explained in the Student Conduct Code. Students are expected to conduct themselves in the course (e.g., on discussion boards, email postings) in compliance with the university's regulations regarding civility.

Academic Integrity
Students are expected to comply with all regulations pertaining to academic honesty. For further information, visit Student Conduct and Community Standards, or contact the office of Student Conduct and Mediation at 541-737-3656.

OAR 576-015-0020 (2) Academic or Scholarly Dishonesty:
a) Academic or Scholarly Dishonesty is defined as an act of deception in which a Student seeks to claim credit for the work or effort of another person, or uses unauthorized materials or fabricated information in any academic work or research, either through the Student's own efforts or the efforts of another.
b) It includes:
   i) CHEATING - use or attempted use of unauthorized materials, information or study aids, or an act of deceit by which a Student attempts to misrepresent mastery of academic effort or information. This includes but is not limited to unauthorized copying or collaboration on a test or assignment, using prohibited materials and texts, any misuse of an electronic device, or using any deceptive means to gain academic credit.

ii) FABRICATION - falsification or invention of any information including but not limited to falsifying research, inventing or exaggerating data, or listing incorrect or fictitious references.

iii) ASSISTING - helping another commit an act of academic dishonesty. This includes but is not limited to paying or bribing someone to acquire a test or assignment, changing someone's grades or academic records, taking a test/doing an assignment for someone else by any means, including misuse of an electronic device. It is a violation of Oregon state law to create and offer to sell part or all of an educational assignment to another person (ORS 165.114).

iv) TAMPERING - altering or interfering with evaluation instruments or documents.

v) PLAGIARISM - representing the words or ideas of another person or presenting someone else's words, ideas, artistry or data as one's own, or using one's own previously submitted work. Plagiarism includes but is not limited to copying another person's work (including unpublished material) without appropriate referencing, presenting someone else's opinions and theories as one's own, or working jointly on a project and then submitting it as one's own.

c) Academic Dishonesty cases are handled initially by the academic units, following the process outlined in the University's Academic Dishonesty Report Form, and will also be referred to SCCS for action under these rules.

Tutoring and Writing Assistance

NetTutor is a leading provider of online tutoring and learner support services fully staffed by experienced, trained and monitored tutors. Students connect to live tutors from any computer that has Internet access. NetTutor provides a virtual whiteboard that allows tutors and students to work on problems in a real time environment. They also have an online writing suite where tutors critique and return essays within 24 to 48 hours. Access NetTutor from within your Canvas class by clicking on the Tools button in your course menu.

The Oregon State Online Writing Suite is also available for students enrolled in Ecampus courses.

TurnItIn

Your instructor may ask you to submit one or more of your writings to Turnitin, a plagiarism prevention service. Your assignment content will be checked for potential plagiarism against Internet sources, academic journal articles, and the papers of other OSU students, for common or borrowed content. Turnitin generates a report that highlights any potentially unoriginal text in your paper. The report may be submitted directly to your instructor or your instructor may elect to have you submit initial drafts through Turnitin, and you will receive the report allowing you the opportunity to make adjustments and ensure that all source material has been properly cited. Papers you submit through Turnitin for this or any class will be added to the OSU Turnitin database and may be checked against other OSU paper submissions. You will retain all rights to your written work. For further information, visit Academic Integrity for Students: Turnitin – What is it?

Student Evaluation of Courses

During Fall, Winter, and Spring term The online Student Evaluation of Teaching system opens to students the Wednesday of week 8 and closes the Sunday before Finals Week. Students receive notification, instructions and the link through their ONID. They may also log into the system via Online Services. Course evaluation results are extremely important and used to help improve courses and the hybrid learning experience for future students. Responses are anonymous (unless a student chooses to "sign" their comments, agreeing to relinquish anonymity) and unavailable to instructors until after grades have been posted. The results of scaled questions and signed comments go to both the instructor and their unit head/supervisor. Anonymous (unsigned) comments go to the instructor only.