CAL POLY HUMBOLDT
University Senate

Resolution to Recommend New Bachelors of Science in Data Science
10-21/22-ICC – February 8, 2022 – First Reading

RESOLVED: That the University Senate of Humboldt State University recommends to the Provost that the new Bachelors of Science in Data Science detailed in proposal 21-1441 be approved.

RATIONALE: The proposed Bachelors of Science in Data Science is intended to support students in developing and practicing skills in synthesizing knowledge and applying contemporary statistics, data analysis, and computational science methods to solve social and environmental problems. Through this program, students will engage in the processes of obtaining, wrangling, curating, managing and processing, and exploring data, defining questions, performing analyses and communicating the results. The Department of Mathematics and the ICC believe this major will provide students with knowledge and skills for future careers in data management as it applies to a variety of fields and specialty areas. There are six new courses being proposed to support this program. This program is being proposed as part of the transition to a Polytechnic University.

Bachelors of Science in Data Science Program Learning Outcomes

1. Students will demonstrate computational skills to extract different types and quantities of data from multiple sources and create visualizations and other data products for various audiences;
2. Students will demonstrate statistical knowledge to build mathematical models and ensure the validity of data and its analysis;
3. Students will demonstrate domain knowledge in one or more key areas of application to gain domain specific information from data and its analysis and to communicate insights from that data that support understanding of and solutions for critical problems within the domain;
4. Students will demonstrate contemporary computer-based and data-oriented analytical skills and related ethical considerations to support a broad synthesis of knowledge including contributions from humanities, sciences and applied sciences, traditional ecological knowledges, and other foundational frameworks for understanding; and
5. Students will communicate effectively with a diverse range of audiences.

Bachelors of Science in Data Science Curriculum
Requirements for the major (63 – 69 units)

Prerequisite (0 – 6 units)
Students may demonstrate calculus readiness by achieving an appropriate score on a department administered placement test, by successful completion of a course in precalculus, or by completing one of the following prerequisite course pathways, or their equivalent:

MATH 102 (4) Algebra and Elementary Functions, OR
MATH 101 (3) College Algebra AND MATH 101T (3) Trigonometry

Lower Division (27 units)
CS 111 (4) Computer Science Foundations
DATA 111 (4) Intro to Programming and Comp Thinking for Data Science
MATH 107 (3) Intro Linear Algebra
STAT 109 (4) Introductory Statistics for the Applied Sciences
MATH 109 (4) Calculus I
MATH 110 (4) Calculus II
DATA 271 (4) Data Structures, Wrangling, and Visualization for Data Science

Upper Division Core (17 units)
DATA 311 (3) Applied Data Analysis
DATA 322 (3) Machine Learning for Data Science
CS 325 (4) Database Design*
DATA 422 (4) Advanced Topics in Data Science
DATA 450 (3) Capstone, Data Science

Upper Division Statistics (4 units)
Select one of the following statistics courses:
STAT 323 (4) Probability and Statistics**
STAT 333 (4) Linear Regression
STAT 404 (4) Multivariate Statistics
STAT 406 (4) Sampling Design and Analysis
STAT 410 (4) Modern Statistical Modeling
Or an advisor approved statistics-related course in an area of application
*CS 111 and DATA 111 will satisfy prerequisites for CS 325
**Requires an additional prerequisite of MATH 210 (4 units) Calculus III

Area of Application/Emphasis (15 units)

To complete the Data Science degree, students are expected to gain expertise in an area to which Data Science may be applied. The area of application/emphasis must be an advisor-approved cohesive set of at least 15 units, 9 of which must be at the upper division level, drawn from existing coursework offered at the university. Suggested areas of emphasis include:

- Mathematics. For students who desire access to more technically demanding careers requiring extensive knowledge of mathematics.
- Biological Sciences. For students who wish to use data science to tackle a diverse set of biological questions in areas ranging from medicine to genomics to evolution.
- Business/Economics. For students with career goals that demand specialized business training. This emphasis will help students appreciate how data science methods support business or economic decision-making and can improve products, services, and organizations.
- Energy. For students interested in combining engineering and environmental science with data science. Career paths include engineering consulting firms, state or federal policy agencies, and private energy industry firms.
- Natural Resources and/or Environmental Planning. For students interested in careers as industry representatives, advocates, consultants, analysts and government planners working on natural resource and/or environmental issues.
- Justice. For students interested in careers which critically analyze crime patterns, access to justice, policy, or advocacy.
- Political Science. For students interested in using data science to help predict, explain, or analyze political phenomena and behavior.

Students also have the option, with advisor approval, to complete a minor or certificate program to complete the area of application/emphasis requirement.

Descriptions of New Courses Proposed as Part of the Major

DATA 111 (4) Introduction to Programming and Computational Thinking for Data Science- Intertwines three perspectives: inferential thinking, computational thinking, and real-world relevance. Students explore real data to understand relationships and patterns while learning critical skills in computer programming and basic statistical inference.

DATA 271 (4) Data Wrangling and Visualization- Provides an intensive, hands-on introduction to data wrangling, cleaning and visualization using a contemporary programming language. Learn

the fundamental skills required to acquire, transform, manipulate, and visualize data in a computing environment that fosters reproducibility. The overall goal is to create actionable data from raw sources and then perform exploratory analysis. These steps will include importing data, viewing data diagnostically, identifying outliers, imputing data, cleaning data, calculating basic statistics and creating informative plots.

DATA 311 (4) Applied Data Analysis- This team- and project-based course provides an opportunity to work through the entire data analysis pipeline in an applied context. Three modules cover (1) cleaning and exploring data, (2) visualizing data and (3) prediction, validation and uncertainty estimation. Students will advance written and verbal communication skills by reporting their findings.

DATA 322 (4) Machine Learning for Data Science- Provides a broad introduction to machine learning, datamining, and statistical pattern recognition. Topics include: (i) Supervised learning (ii), Unsupervised learning and (iii) Best practices in machine learning. The course draws from numerous case studies and applications, with a practical rather than theoretical emphasis.

DATA 422 (4) Advanced Data Science Methods and Ethics- Provides a treatment of advanced topics in Data Science. Topics include: (i) Privacy and Ethics, (ii) Multivariate Statistics, (iii) Big Data and Cloud Computing and (iv) Deep Learning. The course will draw from numerous case studies and applications, with a practical rather than theoretical emphasis.

DATA 450 (3) Data Science Capstone- Semester-long, student-led Data Science group project culminating in a written and an oral presentation. Projects explore an area of application. Preparation for professional life after graduation, including graduate school attendance, working as a Data Scientist, or other quantitative careers.