LSC Use Only No:	LSC Action-Date:	UWUCC USE Only No.	UWUCC Action-Date:	Senate Action Date:
		14-107	AP·11/18/14	App 12/2/14

Curriculum Proposal Cover Sheet-University-Wide Undergraduate Curriculum Committee

Contact Person	Email Address
Jianfeng Wang	jwang@iup.edu
Proposing Department/Unit	Phone
Information System and Decision Sciences	357-5750

Check all appropriate lines and complete information as requested. Use a separate cover sheet for each course proposal and for each program proposal.

1. Course Proposals (check all that apply)

____X__ New Course   _____Course Prefix Change   _____Course Deletion

____Course Revision   _____Course Number and/or Title Change   _____Catalog Description Change

	QBUS 450/550 Data Science for Business
Current Course prefix, number and full title	*Proposed course prefix, number and full title, if changing*

2. Additional Course Designations: check if appropriate

    _____ This course is also proposed as a Liberal Studies Course.   _____ Other: (e.g., Women's Studies,
    _____ This course is also proposed as an Honors College Course.   Pan-African)

3. Program Proposals

_____Catalog Description Change   _____Program Revision

_____New Degree Program   _____Program Title Change   _____Other

_____New Minor Program   _____New Track

Current program name	*Proposed program name, if changing*

4. Approvals

		Date
Department Curriculum Committee Chair(s)	James A Redy 11-11-2014 JAMES A RODGER	11-11-2014
Department Chair(s)	Pankaj PANKAJ	Nov/11/2014 Nov/11/2014
College Curriculum Committee Chair	Pankaj	
College Dean	W C Aargn	11/14/14
Director of Liberal Studies *		
Director of Honors College *		
Provost *		
Additional signatures as appropriate:		
(include title)		
UWUCC Co-Chairs	Gail Sechrist	11/19/14

* where applicable

UNDERGRADUATE AND GRADUATE CURRICULUM AUTHORIZATION FORM

Check As Appropriate:

_____Expedited Review

_____Level I ⇒ New Program □ or Program Revision □

_____Level II ⇒ Degree Program Revision □ or New/Revised COR □

_____Level III Minor Program Revision

_____Application for COR Renewal

_____Deletion Course □ or Track □ or Minor □ or Program □

_____Variability in Program Delivery

_____Program Reactivation

_____New Course ⇒ Enter **CIP Code** (Contact Registrar's Office): _____

_____Major Course Revision

_____Minor Course Revision

__X___Dual-Level or Cross Listing

_____Distance Education

Description of Proposal: QBUS 450/550 –Data Science for Business

Department: Information System and Decision Sciences
Author or Contact Person: _Jianfeng Wang (jwang@iup.edu)_

SIGNATURES OF APPROVAL

Academic Integrity Resource Implications Email Address Date

Department Curriculum
Committee Chair: _____ Jan H Kerh 11-11-2014_

Department Chair: _____ (PANKAJ) NOV/11/2014_

TECC Chair: _____

College Curriculum
Committee Chair: _____ (PANKAJ) NOV/11/2014_

College Dean: _____ 11/14/14_

Graduate Dean: _____

Provost: _____

Graduate Com. Chair: _____

COMMENTS:

_____

APPROVE EXPEDITED REVIEW _____

Dean's Associate, SGSR Dean, SGSR Provost UWGC Chairs

If any signatory above wishes to decline expedited review, write "Decline" on the appropriate signature line and return the proposal to the SGSR Dean's Associate.

To sign off on resource implications, confirm that detailed justification of resource sufficiency has been made and approved.

B. Proposal Format and Content

1. <u>Graduate Curriculum Authorization Form</u> (Appendix D)
   Cover page

2. <u>Course Description and Particulars</u>
   a. <u>Attachments</u>
      i. Course Syllabus: (pages 1-5)
      ii. Bibliography: (page 4-5)

   b. <u>Rationale</u>:
      Data is the new oil. Data is a new class of economic asset. Those were the conclusions of the reports issued by the World Economic Forum, at Davos, in January 2011 and January 2012. Research published in 2011, by MIT economists', shows that companies adopting "data-driven decision-making" achieved significant productivity gains, over other firms. In industry, the hottest job, these days, is the Data Scientist. Data scientists combine technical and statistical skills, analytical thinking, and business acumen. The data scientists trained in computer science departments are quite technical, and understand algorithms well, but are generally lacking important skills in problem formulation, evaluation, and analysis. Those trained in math and statistics departments, in addition to those trained in business schools, tend to have underdeveloped technical skills. This course will cover all of these aspects of being a data scientist.

   c. <u>Other courses or programs</u>:
      i. The course content will not overlap with any courses offered by other departments.
      ii. No other department is planning to use this course as a requirement or elective.

   d. <u>Variable credit</u>:
      This course will not be offered as variable credit
   e. <u>Teaching Plans</u>:
      Instruction will be delivered by one instructor, in a computer lab, that has the appropriate tools and software for applying the concepts through hands on work.

   f. <u>Special Topics</u>:
      This course has not been offered as Special Topics.

   g. <u>Comparisons</u>:

      A reasonable number of universities and colleges offer this multidisciplinary course in various forms and modifications. UC Berkley has an online program in Data Science. Carnegie Mellon University's Master of Science in Information Technology program offers training in data analytics and business intelligence. Columbia University and New York University offer courses in data science related areas.

h. Standards:

This is a course in a newly emerging multidisciplinary area which offer hands-on business skills. UC Berkley, Carnegie Mellon University, New York University etc. all recommend this course.

3. Implementation
    a. Resources:
        i. Faculty: We will be able to offer this course within our currently assigned complement.
        ii. Space and Equipment: Classroom space is adequate. The ECB lab 111 is adequately equipped for this course. The Eberly laboratory is adequately equipped for this course; all software is available either free or through existing subscriptions.
        iii. Laboratory Supplies: The ISDS Department has licensed copies of Excel and SQL Server. Python, R and Google BigQuery are also freely available. However, periodic updates will be required to keep up with the technology.
        iv. Library: There is an adequate source of reading material in Stapleton Library.
        v. Travel Funds: No travel funds will be requested.
        vi. Miscellaneous: There are no other resource requirements or limitations.

    b. Frequency:

This course will be offered once during the academic year.

    c. Enrollment:

Approximately 45 students will be accommodated, in a section, of the course.

4. Catalog Description: See the attached syllabus of record.

5. Logistics
    a. START TERM: Fall 2015
    b. A review of the Registrar's website confirms that this course number IFMG 499/599 is available.
    c. A conversation with the Registrar's Office indicated the CIP code is no longer necessary. The Registrar's Office said that I could use the code 999999 on the Graduate Curriculum Authorization Form.
    d. CATALOG TERM: Fall 2014

6. Miscellaneous: None.

QBUS 450 Data Science for Business   (3c-01-3cr)
Prerequisite: QBUS 215.

QBUS 550 Data Science for Business   (3c-01-3cr)
Prerequisite: QBUS 215

I.   Catalog Description

An introduction to the practice of data science with a broad set of data analytic skills based on building real analytic applications on a variety of real data, including big data, unstructured data and textual data. Skills include accessing and transferring data, applying various analytical frameworks, applying methods from machine learning and data mining, conducting large-scale rigorous evaluations with business goals in mind, and visualizing data analytic results.

II.   Course Objectives/Outcomes:

This advanced course in quantitative analysis is concerned with data-driven decision analytic skills. By the end of this class, students will be able to:
1. Extract and cleanse data from business information systems and other sources.
2. Transfer and load data to an appropriate data analytic platform.
3. Understand and configure open source tools for data analytics.
4. Analyze very large data sets using cloud-based tools.
5. Understand basic concepts and modeling skills in data mining.
6. Understand basic concepts and modeling skills in machine learning.
7. Use data-driven decision analytic skills to improve business decision making.

In addition to the above, QBUS550 graduate students will be able to:
8. Explore the data to discover important patterns and trends
9. Build data-driven models and model-based tools
10. Design and conduct data-oriented experiments with modern information systems

III.   Detailed Course Outline

1. Introduction: Data-Analytic Thinking   (3 hrs)
      Data Science, Engineering, and Data-Driven Decision Making
      Data Processing and "Big Data"; Data and Data Science Capability as a Strategic Asset
      Data Mining and Data Science

2. Business Problems and Data Science Solutions   (3 hrs)
      Fundamental concepts: A set of canonical data mining tasks
      Supervised versus unsupervised data mining.
      Implications for Managing the Data Science Team
      Database Querying, Data Warehousing, Regression Analysis
      Machine Learning and Data Mining

3. Predictive Modeling: From Correlation to Supervised Segmentation   (3 hrs)

Fundamental concepts: Identifying informative attributes; Segmenting data by progressive attribute selection; Supervised Segmentation, Supervised Segmentation with Tree-Structured Models; Visualizing Segmentations; Probability Estimation

4. Fitting a Model to Data   (3 hrs)
   Fundamental concepts: Finding "optimal" model parameters based on data; Objective functions; Loss functions; Regression via Mathematical Functions
   Class Probability Estimation and Logistic "Regression"
   Nonlinear Functions, Support Vector Machines, and Neural Networks

5. Overfitting and Its Avoidance   (3 hrs)
   Fundamental concepts: Generalization; Fitting and overfitting; Complexity control.
   Generalization, Overfitting Avoidance and Complexity Control

6. Similarity, Neighbors, and Clusters   (3 hrs)
   Fundamental concepts: Calculating similarity of objects described by data; using similarity for prediction; Clustering as similarity-based segmentation
   Nearest-Neighbor Reasoning, Clustering

7. Decision Analytic Thinking I: What Is a Good Model?   (3 hrs)
   Fundamental concepts: Careful consideration of what is desired from data science results; Expected value as a key evaluation framework
   Generalizing Beyond Classification
   Evaluation, Baseline Performance, and Implications for Investments in Data

8. Visualizing Model Performance   (3 hrs)
   Fundamental concepts: Visualization of model performance under various kinds of uncertainty; further consideration of what is desired from data mining results.

9. Evidence and Probabilities   (3 hrs)
   Fundamental concepts: Explicit evidence combination with Bayes' Rule; Probabilistic reasoning via assumptions of conditional independence; A Model of Evidence ``Lift"

10. Representing and Mining Text   (3 hrs)
    Fundamental concepts: The importance of constructing mining-friendly data representations; Representation of text for data mining
    Representation, Bag of Words, Term Frequency
    Measuring Sparseness: Inverse Document Frequency

11. Decision Analytic Thinking II: Toward Analytical Engineering   (3 hrs)
    Fundamental concepts: Solving business problems with data science starts with analytical engineering: designing an analytical solution, based on the data, tools, and techniques available.

12. Other Data Science Tasks and Techniques   (3 hrs)
    Fundamental concepts: Our fundamental concepts as the basis of many common data science techniques; the importance of familiarity with the building blocks of data science.
    Profiling: Finding Typical Behavior
    Link Prediction and Social Recommendation
    Data Reduction, Latent Information, and Movie Recommendation
    Bias, Variance, and Ensemble Methods
    Data-Driven Causal Explanation in marketing analysis

13. Data Science and Business Strategy   (3 hrs)

Fundamental concepts: Our principles as the basis of success for a data-driven business;
Acquiring and sustaining competitive advantage via data science; the importance of careful
curation of data science capability.
Thinking Data-Analytically, Reduction
Achieving Competitive Advantage with Data Science
Sustaining Competitive Advantage with Data Science

14. Conclusion   (3 hrs)
Fundamental concepts of Data Science
Applying Fundamental Concepts to a New Problem: Mining Mobile Device Data
Changing the Way We Think about Solutions to Business Problems
Privacy, Ethics, and Mining Data about Individuals

15. Final Exam (only for undergraduates)   (2 hrs)

Total Hours: 44

IV.   Evaluation Methods

Undergraduate Students

Homework	20%
Take Home assessment	20%
Term Project	20%
Final Exam	30%
Participation and Class Contribution: attendance and whether one is active in discussion	10%

Graduate Students

Homework	20%
Take Home assessment	20%
Term Project	20%
Research Paper addressing the construction of data-driven model and design of data-oriented experiments.	30%
Participation and Class Contribution: attendance and whether one is active in discussion	10%

Undergraduate Grading Scale:

A: 90% and above; B: 80% and above but less than 90%; C: 70% and above but less than 80%; D: 60% and above but less than 70%; F: Below 60%

Graduate Grading Scale

A: 90% and above; B: 80% and above but less than 90%; C: 70% and above but less than 80%; F: Below 70%

V.   Course Attendance Policy

In accordance with University policy, individual faculty will denote an attendance policy on specific course syllabi.

VI.   Required Textbook(s), Supplemental Books and Readings

Data Science for Business, by Provost and Fawcett, ISBN-10: 1449361323, O'Reilly Media; 1 edition (August 19, 2013)

3

Doing Data Science, by O'Neil and Schutt, ISBN-10: 1449358659, O'Reilly Media; 1 edition (November 3, 2013)

VII.   Special Resources

No special resource requirements. All the software needed is provided free or already subscribed by the ISDS department. This software is installed in the ISDS lab which is used for the course.

VIII.   Bibliography

Brynjolfsson, E., Hitt, L. M., and Kim, H. H. (2011). Strength in numbers: How does data-driven decision making affect firm performance? Tech. rep., available at SSRN: http://ssrn.com/abstract=1819486.

Business Insider (2012). The Digital 100: The world's most valuable private tech companies. http://www..businessinsider.com/2012-digital-100.

Ciccarelli, F. D., Doerks, T., Von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science*, 311 (5765), 1283-1287.

Lapointe, F. J., and Legendre, P. (1994). A classification of pure mat Scotch whiskies. Applied Statistics, 43 (1), 237-257.

Leigh, D. (1995). Neural networks for credit scoring. In Goonatilake, S., and Treleaven, P. (Eds.), *Intelligent Systems for Finance and Business*, pp. 61-69. John Wiley and Sons, Hoboken, NJ.

Letunic, and Bork (2006). Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23 (1).

Lin, J. H., and Vitter, J. S. (1994). A theory for memory-based learning. *Machine Learning*, 17, 143-167.
Linoff, G. (2008). Data Analysis Using SQL and Excel. Wiley, Indianapolis, IN.

Lloyd, S. P. (1982). Least square quantization in PCM. *IEEE Transactions on Information Theory*, 28 (2), 129-137.\

MacKay, D. (2003). *Information Theory, Inference and Learning Algorithm*ms, Chapter 20. An Example Inference Task: Clustering. Cambridge University Press.

Minelli, M., Chambers, M, and Dhiraj, A. (2013). Big Data, Big Analytics. John Wiley & Sons, Hoboken, NJ.

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.

Swets, J. (1996). *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics*: Collected Papers. Lawrence Erlbaum Associates, Mahwah, NJ.

Swets, J., Dawes, R. M., and Monahan, J. (2000). Better decisioins through science. Scientific American, 283, 82-87.

Tambe, P. (2013). Big Data Investment, Skills, and Firm Value. Working paper, NYU Stern. Available at http://papers.ssrn.com/sol3/papers.cfm? Abstract_id=2294077

Tigani, J. and Naidu, S. (2014). Google BigQuery Analytics. Wiley, Indianapolis, IN.
WEKA (2001). Weka machine learning software. Available:
http://www.cs.waikato.ac.nz/~ml/index.html.

McKinney, W. (2013). Python for Data Analysis. O'Reilly Media, Sebastopol, Ca.

Wikipedia (2012). Determining the number of clusters in a data set. *Wikipedia.*

COURSE ANALYSIS QUESTIONNAIRE

Section A: Details of the Course

A1   How does this course fit into the programs of the department? For what students is the course designed? (Majors, students in other majors, liberal studies). Explain why this content cannot be incorporated into an existing course.

Data sciences for business are necessary for students to grasp fundamental and advanced skills of business analysis, in solving business problems. Students will program throughout the course and are expected to have some programming experience coming in. The emphasis of the course will be on rigor and practical usefulness. The College Corporate Advisory Board members have suggested the need for extensive exposure, of the students to these areas. This course will fit into the programs of the department by meeting this need. The course is designed for our MBA and general business students with necessary background

A2   Does this course require changes in the content of existing courses or requirements for a program? If catalog descriptions of other courses or department programs must be changed as a result of the adoption of this course, please submit as separate proposals all other changes in courses and/or program requirements.

No changes in other current courses or programs, in the department, are required.

A3   Has this course ever been offered at IUP on a trial basis (e.g. as a special topic)? If so, explain the details of the offering (semester/year and number of students).

Part of this course is offered at the special seminar, in fall 2014, in IFMG481/581, by Dr. Wang.

A4   Is this course to be a dual-level course? If so, please note that the graduate approval occurs after the undergraduate.

The course is intended to be listed as dual level.

A5   If this course may be taken for variable credit, what criteria will be used to relate the credits to the learning experience of each student? Who will make this determination and by what procedures?

The course may not be taken for variable credit.

A6   Do other higher education institutions currently offer this course? If so, please list examples (institution, course title).

Similar courses are being offered, in a wide variety of names, in data science related programs, including MBA, MS in Business Analytics, MS in Data Science, Executive Education, Business Undergraduate, and Arts & Science Undergraduate.

Aalto University
Auburn University
Bellarmine University
Buenos Aires University
Brandeis University
Cambridge University
Carnegie Mellon University
Copenhagen Business School
Dalhousie University
Drew University
Duke University
Eindhoven University of Technology
Elon University
Freie Universität Berlin
George Mason University
Georgia Tech
The Hong Kong University of Science and Technology
Indian Institute of Management, Lucknow
INSEAD
Iowa State University
Katholieke Universiteit Leuven
Korea Advanced Institute of Science and Technology
Loras College
MIT Sloan School of Management
Muhlenberg College
New York University
Northwestern University
Notre Dame
Point Park University
Rutgers University
San Jose State University
Seattle Pacific University
Stevens Institute of Technology
Technische Universiteit Eindhoven
Tel Aviv University
Trinity University
Universidad Austral (Argentina)
University of Antwerp
Université Libre de Bruxelles
University of Chicago
University College London
University of Haifa
University of Iowa
University of Minnesota - Twin Cities
University of Minnesota - Duluth
University of North Carolina at Charlotte
University of Pennsylvania (Wharton School)
University of South Florida
University of Southampton

University of Texas at Austin
University of Toronto
University of Washington
Yeshiva University

A7   Is the content, or are the skills, of the proposed course recommended or required by a professional society, accrediting authority, law or other external agency? If so, please provide documentation.

Quite a number of universities and colleges offer this integrated course, in various forms and modifications. For example, in western Pennsylvania, both Carnegie Mellon and Point Park University offer this course.

Section B: Interdisciplinary Implications

B1   Will this course be taught by instructors from more than one department? If so, explain the teaching plan, its rationale, and how the team will adhere to the syllabus of record.

The course is designed to be taught by one ISDS instructor per semester, for MBA and general business majors with the necessary background.

B2   What is the relationship between the content of this course and the content of courses offered by other departments? Summarize your discussions (with other departments) concerning the proposed changes and indicate how any conflicts have been resolved. Please attach relevant memoranda from these departments that clarify their attitudes toward the proposed change(s).

The course does not overlap with any other courses at this University. Although other departments may offer courses with similar topics, this course is specifically designed for the needs, interests, and context required for advanced business analysis, for our MBA and general business students, with the necessary background. This syllabus, of record, proposes the tools and methods used to implement the content of the course.

B3   Will this course be cross-listed with other departments? If so, please summarize the department representatives' discussions concerning the course and indicate how consistency will be maintained across departments.

This course will not be cross listed.

B4   Will seats in this course be made available to students in the School of Continuing Education?

Seats will be made available to Continuing Education students meeting the prerequisites.

Section C: Implementation

C1   Are faculty resources adequate? If you are not requesting or have not been authorized to hire additional faculty, demonstrate how this course will fit into the schedule(s) of current faculty. What will be taught less frequently or in fewer sections to make this possible? Please specify how preparation and equated workload will be assigned for this course.

Faculty resources are currently adequate.

C2   What other resources will be needed to teach this course and how adequate are the current resources? If not adequate, what plans exist for achieving adequacy? Reply in terms of the

following:

Space: Classroom space is adequate. The ECB lab 111 is adequately equipped for this course.
Equipment: The Eberly laboratory is adequately equipped for this course; all software is available, either free or through existing subscriptions.

Laboratory Supplies and other Consumable Goods: The ISDS Department has licensed copies of, Excel, SQL Server and other application software for projects. R and Python are open source tools and already installed in the lab 111. Google BigQuery and Apache Impala are cloud-based applications that are free for educational purposes. However, periodic updates will be required to keep up with the technology.

Library Materials: There is an adequate source of reading material in Stapleton Library.

Travel Funds: No travel funds are needed.

C3   Are any of the resources for this course funded by a grant? If so, what provisions have been made to continue support for this course once the grant has expired? (Attach letters of support from Dean, Provost, etc.)

No resources, for this course, are funded by a grant.

C4   How frequently do you expect this course to be offered? Is this course particularly designed for or restricted to certain seasonal semesters?

Once in an academic year. The course is typically offered in the fall semester.

C5   How many sections of this course do you anticipate offering in any single semester?

At least one section.

C6   How many students do you plan to accommodate in a section of this course? What is the justification for this planned number of students?

Approximately 45 students will be accommodated, in a section, of the course.

C7   Does any professional society recommend enrollment limits or parameters for a course of this nature? If they do, please quote from the appropriate documents.

No professional society recommends enrollment limits or parameters for this course.

C8   If this course is a distance education course, see the Implementation of Distance Education Agreement and the Undergraduate Distance Education Review Form in Appendix D and respond to the questions listed.

Presently, this course is not a distance education course, but a D2L site has been developed.

Section D: Miscellaneous
None.