

# REQUEST FOR NEW COURSE

## 1. General Information.

- a. Submitted by the College of: Gatton College of Business & Economics Today's Date: 02/20/2010
- b. Department/Division: Analytics (formerly Decision Science & Information Systems)
- c. Contact person name: Ram Pakath Email: pakath@uky.edu Phone: 7-4319
- d. Requested Effective Date:  Semester following approval OR  Specific Term/Year<sup>1</sup>: Fall 2010

## 2. Designation and Description of Proposed Course.

- a. Prefix and Number: AN 420G
- b. Full Title: Data Mining
- c. Transcript Title (if full title is more than 40 characters): Data Mining
- d. To be Cross-Listed<sup>2</sup> with (Prefix and Number): \_\_\_\_\_
- e. Courses must be described by at least one of the meeting patterns below. Include number of actual contact hours<sup>3</sup> for each meeting pattern type.
- |             |          |                         |            |            |                         |            |          |              |           |
|-------------|----------|-------------------------|------------|------------|-------------------------|------------|----------|--------------|-----------|
| 3.0 Lecture | _____    | Laboratory <sup>1</sup> | _____      | Recitation | _____                   | Discussion | _____    | Indep. Study |           |
| _____       | Clinical | _____                   | Colloquium | _____      | Practicum               | _____      | Research | _____        | Residency |
| _____       | Seminar  | _____                   | Studio     | _____      | Other – Please explain: | _____      |          |              |           |
- f. Identify a grading system:  Letter (A, B, C, etc.)  Pass/Fail
- g. Number of credits: 3.0
- h. Is this course repeatable for additional credit? YES  NO
- If YES: Maximum number of credit hours: \_\_\_\_\_
- If YES: Will this course allow multiple registrations during the same semester? YES  NO

## i. Course Description for Bulletin:

Data mining is concerned with tools and techniques to help a data/business analyst numerically and visually explore vast data sets, classify data, predict outcomes, or identify associations, patterns, and exceptional events. In practical terms, such capabilities allow firms to better segment markets, evaluate and classify stocks, identify prospective customers, foretell contingencies and catastrophes, identify defaulters and fraudulent transactions, measure churn, identify threats, perform service requests, bundle goods and services, recognize how events (e.g., purchases) are likely to unfold over time, and so on. Such capabilities often make the difference between survival and demise in today's, increasingly global, increasingly competitive, and increasingly volatile business settings.

## j. Prerequisites, if any: Completion of all college pre-major requirements and admission to Upper Division or \_\_\_\_\_

<sup>1</sup> Courses are typically made effective for the semester following approval. No course will be made effective until all approvals are received.

<sup>2</sup> The chair of the cross-listing department must sign off on the Signature Routing Log.

<sup>3</sup> In general, undergraduate courses are developed on the principle that one semester hour of credit represents one hour of classroom meeting per week for a semester, exclusive of any laboratory meeting. Laboratory meeting, generally, represents at least two hours per week for a semester for one credit hour. (from SR 5.2.1)

# REQUEST FOR NEW COURSE

Graduate student status in Business & Economics. Non B&E Upper Division undergraduate students and Graduate students may be enrolled with the consent of the instructor.

- k. Will this course also be offered through Distance Learning? YES<sup>4</sup>  NO
- l. Supplementary teaching component, if any:  Community-Based Experience  Service Learning  Both
3. Will this course be taught off campus? YES  NO
4. Frequency of Course Offering.
- a. Course will be offered (check all that apply):  Fall  Spring  Summer
- b. Will the course be offered every year? YES  NO   
If NO, explain: \_\_\_\_\_
5. Are facilities and personnel necessary for the proposed new course available? YES  NO   
If NO, explain: \_\_\_\_\_
6. What enrollment (per section per semester) may reasonably be expected? 15-20
7. Anticipated Student Demand.
- a. Will this course serve students primarily within the degree program? YES  NO
- b. Will it be of interest to a significant number of students outside the degree pgm? YES  NO   
If YES, explain: Should be of interest to any major concerned with gathering insights from available, opportunistic, secondary data. Likely programs include Economics, Accounting, Statistics, Computer Science, and others.
8. Check the category most applicable to this course:
- Traditional -- Offered in Corresponding Departments at Universities Elsewhere
- Relatively New -- Now Being Widely Established
- Not Yet Found in Many (or Any) Other Universities
9. Course Relationship to Program(s).
- a. Is this course part of a proposed new program? YES  NO   
If YES, name the proposed new program: Analytics (formerly, Decision Science and Information Systems)
- b. Will this course be a new requirement<sup>5</sup> for ANY program? YES  NO   
If YES<sup>5</sup>, list affected programs: Analytics (formerly, Decision Science and Information Systems)
10. Information to be Placed on Syllabus.
- a. Is the course 400G or 500? YES  NO   
If YES, the *differentiation for undergraduate and graduate students must be included* in the information required in 10.b. You must include: (i) identification of additional assignments by the graduate students; and/or (ii) establishment of different grading criteria in the course for graduate students. (See SR 3.1.4.)

<sup>4</sup> You must *also* submit the Distance Learning Form in order for the proposed course to be considered for DL delivery.

<sup>5</sup> In order to change a program, a program change form must also be submitted.

## REQUEST FOR NEW COURSE

- b.  The syllabus, including course description, student learning outcomes, and grading policies (and 400G-/500-level grading differentiation if applicable, from **10.a** above) are attached.

# REQUEST FOR NEW COURSE

## Signature Routing Log

**General Information:**

Course Prefix and Number: AN 420G

Proposal Contact Person Name: Ram Pakath      Phone: 7-4319      Email: pakath@uky.edu

**INSTRUCTIONS:**

Identify the groups or individuals reviewing the proposal; note the date of approval; offer a contact person for each entry; and obtain signature of person authorized to report approval.

**Internal College Approvals and Course Cross-listing Approvals:**

Reviewing Group	Date Approved	Contact Person (name/phone/email)	Signature
School of Mgmt	2/10	Scott Kelley 7 134251 @uky.edu	Scott W Kelley
Undergrad Studies	4/23/10	Nancy Johnson 7 129761 nbj@uky.edu	Nancy Johnson
Gatton Faculty	4/30/10	Merrill Hackbart 7 13592 @uky.edu	Merrill Hackbart
		1 1	
		1 1	

**External-to-College Approvals:**

Council	Date Approved	Signature	Approval of Revision <sup>6</sup>
Undergraduate Council			
Graduate Council			
Health Care Colleges Council			
Senate Council Approval		University Senate Approval	

**Comments:**

<sup>6</sup> Councils use this space to indicate approval of revisions made subsequent to that council's approval, if deemed necessary by the revising council.

**AN 420G-001: DATA MINING****Departmental Prefix – DSIS; College Prefix – BE****T, R – 11:00 AM – 12:15 PM : Room: BE 105****Instructor: Dr. Ram Pakath****Office – BE 425 D : Office Hours – T, R – 1:00 – 2:00 PM or By Appointment  
Phone: 859-257-4319; Email – [pakath@uky.edu](mailto:pakath@uky.edu) (preferred contact mode)****Course Description:**

Data mining is concerned with tools and techniques to help a data/business analyst numerically and visually explore vast data sets, classify data, predict outcomes, or identify associations, patterns, and exceptional events. In practical terms, such capabilities allow firms to better segment markets, evaluate and classify stocks, identify prospective customers, foretell contingencies and catastrophes, identify defaulters and fraudulent transactions, measure churn, identify threats, perform service requests, bundle goods and services, recognize how events (e.g., purchases) are likely to unfold over time, and so on. Such capabilities often make the difference between survival and demise in today's, increasingly global, increasingly competitive, and increasingly volatile business settings.

This course is an introduction to the rapidly blossoming field of Data Mining, a corner stone of Business Analytics at several Fortune 500 firms. It is one of twelve Ahead-of-the-Curve Careers identified by US News, March 24-31, 2008. A recent (February 23, 2009) Global Knowledge report, cites Business Intelligence, of which Business Analytics is a part, as one example of a "hot" IT specialty in current times. A March 2, 2009 Business Week Special Report showcases well-known US corporations' use of Business Intelligence software to cut costs, manage risks, and increase profits.

**Course Goals:**

To expose students to key issues in data preparation, dimension reduction and partitioning, generating descriptive statistics, creating visual representations of data, and building classifier models using techniques such as the Naïve Classifier, the Naïve Bayes Classifier, the Nearest Neighbor Classifier, Classification Trees, and Logistic Regression. We will also examine various Classifier Evaluation approaches. If time allows it, students will also be exposed to Artificial Neural Nets, Association Rule Mining, and Clustering techniques.

**Intended Student Learning Outcomes:**

At the end of the course, students should:-

- Understand how data mining differs from, and has overlaps with, traditional statistical analysis and information retrieval/querying.
- Understand the difference between explanatory modeling and predictive modeling.
- Understand how to conduct descriptive and visual analysis of data using spreadsheets and specialized software.
- Understand pre-processing steps including data cleaning, data transformation, and data partitioning.
- Understand predominant classification and prediction methods in data mining.

- Understand how to set up and execute particular data mining models using specialized software.
- Understand how to interpret data mining output reports including confusion matrices, error matrices, lift charts, decile-wise lift charts, etc.
- Understand how to evaluate and choose between multiple models for a given decision task.
- Understand the strengths and weaknesses of data mining as an analytical tool

**Course Pre-requisites:**

(1) Completion of all College Pre-major Requirements (2) Admission to Upper Division or Graduate Student status in Business & Economics (3) Non B&E Upper Division Undergraduate Students and Graduate Students may be enrolled with Consent of the Instructor.

**Course Text:**

Data Mining for Business Intelligence: Concepts Techniques, and Applications in Microsoft Office Excel with XLMiner, G. Shmueli, N. R. Patel, and P. C. Bruce, Wiley, 2007, 0-470-08485-5

Note: The book comes bundled with a 6-month license for the XLMiner software package in the inside back cover. Check to make sure the pouch on the inside back cover is sealed and has a card in it.

We will use appropriate versions of following software in this class on an as-needed basis – MS Excel, Tibco Spotfire, XLMiner, WEKA, and SAS Enterprise Miner.

**Recommended Additional Readings:**

**Text References:**

1. Introduction to Data Mining, P. Tan, M. Steinbach, and V. Kumar, Addison Wesley, 2006 edition.
2. Applied Data Mining – Statistical Methods for Business & Industry, P. Giudici, Wiley, 2003 edition (reprinted 2007)
3. Data Mining – Concepts, Models, methods, and Algorithms, M. Kantardzic, Wiley Inter-Science, 2003 edition.
4. Making Sense of Data – A Practical guide to Exploratory Data Analysis and Data Mining, G. J. Myatt, Wiley Inter-Science, 2007 edition.

**Non-text References:-**

1. Competing on Analytics – The New Science of Winning, T. H. Davenport and J. G. Harris, Harvard Business School Press, 2007.
2. Freakonomics – A Rogue Economist Explores the Hidden Side of Everything, S. D. Levitt and S. J. Dubner, William Morrow/Harper Collins, 2006.
3. Super Crunchers – Why Thinking by Numbers is the New Way to be Smart, I. Ayres, Bantam Books, 2007.

I may draw upon other references (both books and articles), such as those above, on an as-needed basis.

**Student Evaluation Criteria:**

Problem Assignments:	50 points	Periodic; TBA
In-class/Take-Home Quizzes/Exams:	30 points	Periodic; TBA
Case Study *	20 points	TBA.
Theory/Application Paper (Grad Students only)**	25 points	TBA

\* Depending on enrollment size, individual students or student teams will be assigned data mining cases from the text or other sources. As we progress through material in class, students will work in parallel on their cases, using applicable knowledge from what was taught. Toward the end of the semester, each student will prepare a Final Report detailing his/her assigned case, analysis, and recommendations made. I will also ask for an Interim (progress) Report from each of you around the middle or three-quarters of the way into the semester.

\*\* A paper-writing assignment is required of Graduate Students only. I will discuss potential topics and paper specifications with individual students early in the semester.

**Course Grading Scheme:**

The default grading scheme for Undergraduate students is:

A = [90 - 100]; B = [80 - 90]; C = [70 - 80]; D = [60 - 70]; E = [00 - 60)

The default grading scheme for Graduate students is:

A = [90 - 100]; B = [80 - 90]; C = [70 - 80]; E = [00 - 70)

Note 1: During the semester, I shall use a 5-point scale, as shown above for undergraduate students, to assign grades to all students. However, graduate students earning a final, cumulative course score that translates to a "D," will be awarded an "E" for the course as the Grad School does not permit D grades.

Note 2: I may choose to relax the above default grading schemes, at my discretion.

Note 3: All undergraduate students will be provided with a Midterm Evaluation of course performance during the midterm grades submission window of the college, based on criteria in the syllabus.

**Tentative Course Schedule:**

Time and circumstances permitting, I hope to address material pertaining to several of the following Data Mining topics during this semester.

[Note:- Dates for assignment submission, exams, project progress reports, final project reports, and the theory/application paper (only for Graduate students) will be determined closer to the beginning of the Fall 2010 term]:

1. Introduction & Preliminaries
2. Basics of Data Base Design and Query-by-Example with MS Access 2007 -- Optional
3. Data Exploration & Visualization with XLMiner 3, MS Excel 2003/2007, and TIBCO Spotfire
4. Data Dimension Reduction & Partitioning
5. The Naïve Classifier & the k-Nearest Neighbor Classifier
6. The Naïve Bayes Classifier
7. Classification with Classification Trees
8. Logistic Regression
9. Evaluating Classifier Performance
10. Hierarchical (Agglomerative) Clustering
11. k-Means Clustering
12. Neural Networks
13. Association Rule Mining

### Course Policies:-

As a general rule, all students must adhere to Student Rights & Responsibilities-related information available in the Rules of the University Senate. I elaborate below, on some of the more important aspects of student responsibility for the purposes of this class.

1. **Disabilities Accommodation:-** If you have a documented disability that requires academic accommodations, please see me as soon as possible during scheduled office hours. In order to receive accommodations in this course, you must provide me with a Letter of Accommodation from the Disability Resource Center (Room 2, Alumni Gym, 257-2754, email address [jkarnes@email.uky.edu](mailto:jkarnes@email.uky.edu)) for coordination of campus disability services available to students with disabilities.
2. **Attendance:-** Any curve that I may choose to apply to your grade at any point will be some function of your attendance during all or part of the preceding weeks. Please make sure that you answer when your name is called, usually within the first 5 minutes of class.

**Excused Absences & Verification:-** If your absence must be excused, please make sure it is with valid cause as noted in Section 5.2.4.2—Excused Absences in the Rules of the University Senate. This excused absence policy is especially critical with regard to any homework submission and examination dates.

I will need adequate proof in the form of written documentation from an appropriate authoritative source or sources. Such proof must be submitted before the absence when possible/so noted and no later than one week following an absence, under any circumstances, for the absence to be excused.

3. **Submission of Assignments:-** Homework must usually be turned in on paper, and sometimes via email, and must contain the Homework Number, your full name (First, MI, Last), and the date of submission. Arrange your sheets in ascending sequence of problem number. If a hardcopy submission, staple all sheets at the top left-hand corner. All homework must be turned in on the date and at the time due, personally, by each student. Do not send your homework in with anybody else, or fax it, or email it (unless you are asked to) or hand over directly to my TAs (if any). Late submissions will not be accepted unless you have a valid, verifiable excuse, and will be awarded a grade of zero
4. **Makeup Opportunities:-** If you miss a scheduled exam/quiz with valid excuse, I will offer you one, common makeup opportunity along with all students who missed it. Please avail of this one opportunity or accept a grade of zero for the exam/quiz. If you do have an excused absence when homework or



other assignment is due, personally turn in your work to me the very next day following your excused absence period. If a hardcopy submission, and I am not in, slide your work under my office door with a note indicating date and time turned in or have the DSIS Area Secretary (Ms. Karen Harmon in Suite 425 B&E) make this note for you and accept the work on my behalf. Any homework submitted without regard to these guidelines will receive a grade of zero.

#### 5. **Communication Policy:-**

- a. Make it a point to stay in touch with your email account at least twice each day (morning and night). Make sure that the account you provide is in good working order (i.e., has sufficient inbox room; is active (not suspended), etc.) throughout the semester. Should your email address change for any reason, immediately send me a brief message to this effect (also stating your full name).
- b. I have established a UKClass folder (accessible from any restricted-access campus micro lab) for our purposes. Here is how you can access your folder. Click on "UKClass" on the desktop. Select "B&E" and then, "AN420G." Within "AN420G", look for other folders related to this class. I will post the syllabus, my lecture notes, important announcements, etc., in these folders. As with your email accounts, stay in touch with your UKClass folders continually.
- c. If you wish to have a personal meeting, please stop by during my office hours (T, R: 1:00-2:00 PM). Please do not walk in arbitrarily as I do have other demands on my time.
- d. If my office hours are not convenient, first try resolving the issue through an email.
- e. If email discussions will not work for you, set up an appointment (either in-class, via email, or via telephone) for a personal meeting outside of walk-in (i.e., office) hours.
- f. The least preferred mode of interaction would be a voice mail message (I tend not to check voice mail as frequently as I check email, so my response may be slower or not forthcoming. Further, I prefer not to play telephone tag.)

#### 6. **Classroom Conduct Policy:-**

- a. (a) Do not walk out before I dismiss class or wander in and out of class for any reason (i.e., finish trips to the rest room, the vending machines, the water fountain, etc., before you come in). Please understand that late arrivals, early departures, and other movement are disruptive to the group and the teacher. Likewise, reserve chats, jokes, newspaper reading, laying your head down/sleeping, cell phone usage (shut them off while in class), and such, for outside of class. Remember, you always have the option of staying away from class in case you're in the mood for something other than what's going on in class – feel free to exercise that option.
- b. This class is being held in a computer lab. During class, these computers in the lab are intended strictly for class-related use and as directed by the instructor and his/her assistants. Do not misuse the computers during class.
- c. The penalty for any conduct violation could be an absence for the day in question at the minimum or as severe as an E in the course, depending on circumstances.
- d. What if you suddenly fall ill or suffer some other form of distress during class? If your condition permits it, quietly pack up your belongings, write me a brief note, leave the note on my desk, and depart. If you are unable to write, just pack up your belongings and quietly leave. In either case, contact me with necessary details before the next class meeting or as soon as you are well enough to.

#### 7. **Academic Integrity:-**

- a. Plagiarism includes reproducing someone else's work, whether it be published article, chapter of a book, a paper from a friend or some file, or whatever. Plagiarism also includes the practice of employing or allowing another person to alter or revise the work which a student submits as

his/her own, whoever that other person may be. Students may discuss assignments among themselves or with an instructor or tutor, but when the actual work is done, it must be done by the student, and the student alone.

- b. Cheating is defined by its general usage. It includes, but is not limited to, the wrongfully giving, taking, or presenting any information or material by a student with the intent of aiding himself/herself or another on any academic work which is considered in any way in the determination of the final grade. The fact that a student could not have benefited from an action is not by itself proof that the action does not constitute cheating.
- c. The minimum penalty is a "0" for the activity in question. However, significantly more stringent sanctions are possible. Please refer to Section 6.3.0—Academic Offenses and Procedures in the Rules of the University Senate.

8. **Group work & Student Collaboration:-** In activities that call for student collaboration, all students involved in a group effort are expected to contribute equally to the final outcome. In the event there are verifiable complaints about a group member from others in the group, the student in question will be awarded a grade of zero for the activity.