**FLORIDA ATLANTIC UNIVERSITY**

Graduate Programs—NEW COURSE PROPOSAL

<table>
<thead>
<tr>
<th>Department: Department of Computer and Electrical Engineering and Computer Science</th>
<th>College: College of Engineering and Computer Science</th>
</tr>
</thead>
</table>

**Recommended Course Identification:**
Prefix: CAP Course Number: 6771 Lab Code (L or C): 
(TO OBTAIN A COURSE NUMBER, CONTACT: Maldonado@fau.edu)

**Complete Course Title:** Data Mining for Bioinformatics

**Credits:** 3

**Textbook Information:** Data Mining: Practical Machine Learning Tools and Techniques, by I. H. Witten and E. Frank (3rd Edition); Selected Articles and Papers.

**Grading (Select only one grading option):** Regular: X Satisfactory/Unsatisfactory: 

**Course Description:** No more than three lines: This course deals with the principals of data mining as it relates to bioinformatics. Topics covered include gene selection, class imbalance, classification, biomarker discovery, and prediction models. No prior knowledge of biology is required.

**Prerequisites:** Graduate level status or permission of the instructor

**Corequisites:** N/A

**Registration Controls (Major, College, Level):**

---

*Prerequisites, corequisites and registration controls will be enforced for all course sections.

**Minimum Qualifications needed to teach this course:**
Member of the graduate faculty of FAU and has a terminal degree in the subject area (or a closely related field)

Faculty contact, email and complete phone number:
Taghi M. Khosghoifar, khosgho@fau.edu
561-297-3994

**Please consult and list departments that might be affected by the new course and attach comments:**

<table>
<thead>
<tr>
<th>Approved by:</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department Chair: [Signature]</td>
<td>05/15/15</td>
</tr>
<tr>
<td>College Curriculum Chair: [Signature]</td>
<td>05/15/15</td>
</tr>
<tr>
<td>College Dean: [Signature]</td>
<td>05/15/15</td>
</tr>
<tr>
<td>UGPC Chair: [Signature]</td>
<td>05/15/15</td>
</tr>
<tr>
<td>Graduate College Dean: [Signature]</td>
<td>05/15/15</td>
</tr>
<tr>
<td>UFS President: [Signature]</td>
<td>05/15/15</td>
</tr>
<tr>
<td>Provost: [Signature]</td>
<td>05/15/15</td>
</tr>
</tbody>
</table>

---


3. Consent from affected departments (attach if necessary)

Email this form and syllabus to UGPC@fau.edu one week before the University Graduate Programs Committee meeting so that materials may be viewed on the UGPC website prior to the meeting.

FAUnwvsxGrad Revised November 2014
1. **Course title/number, number of credit hours**

<table>
<thead>
<tr>
<th>Data Mining for Bioinformatics</th>
<th>3 credit hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAP 6771</td>
<td></td>
</tr>
</tbody>
</table>

2. **Course prerequisites, corequisites, and where the course fits in the program of study**

Prerequisites: Graduate standing or permission of instructor

3. **Course logistics**

*Term: Fall 2016*

This is a classroom lecture course with DL sections.

*Class location and time: TBA*

4. **Instructor contact information**

<table>
<thead>
<tr>
<th>Instructor’s name</th>
<th>Dr. Taghi M Khoshgoftaar, Professor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office address</td>
<td>Engineering East Bldg., Room 511</td>
</tr>
<tr>
<td>Office Hours</td>
<td>TBA</td>
</tr>
<tr>
<td>Contact telephone number</td>
<td>561-297-3994</td>
</tr>
<tr>
<td>Email address</td>
<td><a href="mailto:khoshgof@fau.edu">khoshgof@fau.edu</a></td>
</tr>
</tbody>
</table>

5. **TA contact information**

6. **Course description**

This course deals with the principals of data mining as it relates to bioinformatics. Topics covered include gene selection, class imbalance, classification, biomarker discovery, and prediction models. No prior knowledge of biology is required.

7. **Course objectives/student learning outcomes/program outcomes**

| Course objectives | Enable students to understand the basic concept of data mining algorithms with an emphasis on their application and utilization on bioinformatics data |

8. **Course evaluation method**

| Assignments (Homework, Programming, etc.) - 45% |
| Term Project – 40% |
| Paper Presentation – 15% |

9. **Course grading scale**

Grading Scale:

10. **Policy on makeup tests, late work, and incompletes**

Assignments are to be submitted on time, with possible point penalties for late submissions. In no case
**11. Special course requirements**

**12. Classroom etiquette policy**

University policy requires that in order to enhance and maintain a productive atmosphere for education, personal communication devices, such as cellular phones, are to be disabled in class sessions, and laptops are only to be used for note taking and related activities.

**13. Disability policy statement**

In compliance with the Americans with Disabilities Act (ADA), students who require special accommodations due to a disability to properly execute coursework must register with the Office for Students with Disabilities (OSD) located in Boca Raton campus, SU 133 (561) 297-3880 and follow all OSD procedures.

**14. Honor code policy**

Students at Florida Atlantic University are expected to maintain the highest ethical standards. Academic dishonesty is considered a serious breach of these ethical standards, because it interferes with the University mission to provide a high quality education in which no student enjoys unfair advantage over any other. Academic dishonesty is also destructive of the university community, which is grounded in a system of mutual trust and place high value on personal integrity and individual responsibility. Harsh penalties are associated with academic dishonesty. See University Regulation 4.001 at [www.fau.edu/regulations/chapter4/4.001_Code_of_Academic_Integrity.pdf](http://www.fau.edu/regulations/chapter4/4.001_Code_of_Academic_Integrity.pdf)

**15. Required texts/reading**

2. Selected articles and papers are posted on the course website.

**16. Supplementary/recommended readings**

**17. Course topical outline, including dates for exams/ quizzes, papers, completion of reading**
Course Topical Outline

• Introduction to Bioinformatics
  o Basic Genetics
  o Available Online Tools and Databases
  o Details and Challenges Associated with the Bioinformatics Data

• Classification
  o Tumor Diagnosis Models
  o Patient Response Prediction Models
  o Types of Errors
  o Performance Metrics
  o Cost-Sensitive Classifiers

• Ensemble Learning for Bioinformatics
  o Why Ensemble Learning is Beneficial to Bioinformatics
  o Strong and Weak Classifiers
  o Ensemble Vs Cost Sensitive Classifiers
  o Bagging
  o Boosting
  o Random Forest

• Gene Selection
  o Biomarker Identification Through Gene Selection
  o Filter-based Gene Ranking
  o Filter-based Subset Selection
  o Wrapper-based Subset Selections
  o Imbedded Gene Selection
  o Ensemble Gene Selection
  o Gene Selection Stability

• Other Challenges
  o Small Class of Interest/Class Imbalance
  o Source and Effects of Data Noise on Bioinformatics Data