Catalog Description:

Data collected in insurance and financial services have become richer and richer with finer demographic information, aggregated information across different branches and business lines, and other public financial information. Mining this type of big data sets will enable insurance companies to provide personalized services to customers and to better manage their own risks. After reviewing traditional prediction techniques and loss models in insurance such as bankruptcy prediction, predictive loss modeling, stochastic loss reserving, generalized linear modeling, survival regression, neural network analysis and others, this course presents some unique computational and statistical techniques in analyzing big data, and demonstrates how to apply these techniques to solve practical problems in risk management. Some techniques include: using softwares (e.g., Hadoop) for distributed data management and processing, designing scalable algorithms, understanding the statistical issues of heterogeneity, spurious correlation, noise accumulation, incidental endogeneity, etc. in building predictive models and loss models. This course will make extensive use of the open-source R programming language, which is freely available at www.r-project.org.

Course Objectives

- get familiar with predictive techniques in insurance;
- get familiar with loss models in insurance;
- learn new computational and statistical techniques in analyzing big data in insurance
- develop computer programs implementing these models so as to solve practical problems in insurance.

References

Research papers will be used throughout this course although some standard textbooks in predictive modeling, loss models, statistical learning, high frequency data analysis and functional data analysis will be referred to. See course outline below for details.

Grading

The grades will be based on two projects. One is to analyze insurance data by traditional techniques, and another is to address the big data issues with emphasis on i) how big data challenge traditional techniques in insurance, and ii) what benefits big data can provide to insurance companies and customers. Grades will be awarded on a +/- basis, and the following guaranteed scale applies. Grades may be moved upward based on difficulty, but not downward:

<table>
<thead>
<tr>
<th>Grade</th>
<th>A+</th>
<th>A</th>
<th>A-</th>
<th>B+</th>
<th>B</th>
<th>B-</th>
<th>C+</th>
<th>C</th>
<th>C-</th>
<th>D+</th>
<th>D</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>100</td>
<td>95</td>
<td>90</td>
<td>85</td>
<td>80</td>
<td>75</td>
<td>70</td>
<td>65</td>
<td>60</td>
<td>55</td>
<td>50</td>
<td>&lt;50</td>
</tr>
</tbody>
</table>

Attendance

Attendance is not formally taken. However, it is strongly suggested that students do not miss class as most students will have difficulties learning materials without attending the lectures.
Course outline

- **Week 1:** credit scoring and bankruptcy prediction.

- **Weeks 2 and 3:** predictive modeling in insurance. The reference book is 'Predictive Modeling Applications in Actuarial Science' edited by Frees, Derrig and Meyer.

- **Weeks 4 and 5:** loss models in insurance. The reference book is 'Loss Models: Further Topics' by Klugman and Panjer.

- **Week 6:** distributed data management and processing. Hadoop, MapReduce and Cloud computing will be discussed.

- **Weeks 7, 8 and 9:** high dimensional/frequency data analysis. The reference book is 'Statistics for High-Dimensional Data: Methods, Theory and Applications' by Buhlman and Van de Geer.

- **Weeks 10, 11 and 12:** functional time series analysis. The reference book is 'Inference for Functional Data with Applications' by Horváth and Kokoszka.

- **Weeks 13 and 14:** big data. How to design scalable algorithms and how to understand the statistical issues of heterogeneity, spurious correlation, noise accumulation, incidental endogeneity, etc. in building predictive models and loss models.

- **Week 15:** presentation/discussion. The first project is on building predictive and loss models for insurance data, which is due on Week 8. The second project is to understand how big data challenge the traditional techniques and what benefits big data can provide to insurance company and customers. Presentation focuses on the second project.

Remarks

- Students exhibiting disruptive behavior, including talking, sleeping, talking on cell phones or disturbing other students will be asked to leave.

- Please advise the instructor if you have a documented disability that needs to be accommodated.

- As members of the academic community, students are expected to recognize and uphold standards of intellectual and academic integrity. See the University’s policy on Academic Honesty (Section 409, http://www2.gsu.edu/~wwfhb/sec409.html) for details.