

**IS 687**  
**Transaction Mining / Fraud Detection**  
**Course Syllabus**  
**Fall 2014**

**Description of Course:** In recent years, financial fraud, including credit card fraud, corporate fraud and money laundering, has attracted a great deal of concern and attention. Data mining plays an important role in fraud detection and is known as a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and subsequently gain knowledge from data. This course presents methods applicable to transaction mining and fraud detection. Hands-on training will be gained with the Weka data mining workbench. This course will provide the knowledge and skills to analyze data to find risks including detection of terrorists, fraudsters and money launderers.

Instructor: Alex Rudniy

Office: Fenster Hall 461

E-Mail: rudniy@njit.edu

**Office Hours:**

Monday 4:30pm -5:45pm

Tuesday 4:30pm -5:45pm

**Textbook:** Witten, Frank, Hall. Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition. ISBN 0123748569.

**Grading Policy:**

30% Midterm exam (closed notes)

30% Final project (open notes)

30% Paper presentations and assignments

10% Moodle/Class participation

<b>Week</b>	<b>Topic</b>	<b>Textbook Chapter</b>	<b>Papers for Reading and In-Class Presentation (will be posted on Moodle)</b>	<b>Homework</b>
1	Electronic crime. Types of fraud. Introduction to data mining. Supervised vs. unsupervised learning. Industry standards. Benford's Law.	1	<p>The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Ngai, Hu, Wong, Chen, Sun. Decision Support Systems 50 (2011) 559–569.</p> <p>The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. Durtschi, Hillison, Pacini. Journal of Forensic Accounting 1524-5586/Vol.V(2004), pp. 17-34.</p>	Reading as assigned.
2	Concepts. Instances. Attributes. Preparing data. Sparse data, missing values. Weka Data Mining Workbench.	2, 10 (short)	<p>Data Mining for Fraud Detection: Toward an Improvement on Internal Control Systems? Jans, Lybaert, Vanhoof.</p> <p>Automation and Disruption in Stolen Payment Card Markets. Peacock, Friedman. 13th Annual Workshop on the Economics and Inf. Sec., Penn State U, 2014.</p>	1. Install Weka. 2. Exercise 17.1.2 (textbook page 560)
3	Trees. Classification Rules.	3, 4	<p>A model to detect potentially fraudulent/abnormal wires of an insurance company: an unsupervised rule-based approach. Kim, Vasarhlyi. J. of Emerging Technologies in Accounting, Vol 9, 2012, 95-110.</p> <p>A Comparative Analysis of Decision Trees Vis-a-vis Other Computational Data Mining Techniques in Automotive Insurance Fraud Detection. Gepp, Wilson, Kumar, Bhattacharya. Journal of Data Science 10(2012), 537-561.</p>	Exercises 17.1.8, 17.1.9, 17.1.10.

			A cost-sensitive decision tree approach for fraud detection. Sahin, Bulkan, Duman. Expert Systems with Applications 40 (2013) 5916–5923.	
4	Association Rules. K-nearest-neighbor. Clusters.	3, 4	Association rules applied to credit card fraud detection. Sanchez, Vila, Cerda, Serrano. Expert Systems with Applications 36 (2009) 3630–3640.  (Clustering) Mining Information from Plastic Card Transaction Streams. Tasoulis, Adams, Weston, Hand. 18th Int. Conf. on Comp. Stat., Porto, Portugal, 2008.	Ex. 17.2.1, 17.2.2, 17.2.3, 17.2.6, 17.2.8
5	Naïve Bayes. Linear models. Linear Regression, Logistic Regression.	4	(Bayes) Credit Card Fraud Detection: Personalized or Aggregated Model. Alowais, Soon. 2012.  (Regression) Data mining for credit card fraud: A comparative study. Bhattacharyya, Jha, Tharakunnel, Westland. 2011	TBA
6	Midterm			
7	Training, Testing, Cross-Validation. Lift Charts. ROC Curves. Precision-Recall curves	5	ROC Graphs: Notes and Practical Considerations for Data Mining Researchers. Fawcett. 2003.  Measuring classification performance: the hmeasure package. Anagnostopoulos, Hand, Adams, 2012.	TBA
8	Decision Trees in Detail: C4.5. Pruning. From trees to rules. Classification Rules. RIPPER.	6	(C4.5, RIPPER) Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection. Chan, Stolfo. 1998.	TBA
9	Support Vector Machines. Support Vector Regression. Bayesian Networks.	6	(SVM) Detecting Click Fraud in Online Advertising: A Data Mining Approach. Oentaryo et al. 2013.  (Bayes net, neural net) Data Mining techniques for the detection of fraudulent financial statements. Kirkos et al. 2007.	TBA
10	Neural Networks. Perceptrons. Backpropagation.	6	(backpropagation, bayes) Minority Report in Fraud Detection: Classification of Skewed Data. Phua, et al.	TBA

	Expectation Maximization. Semisupervised Learning		2004.  An ANN-based auditor decision support system using Benford's law. Bhattacharya et al. 2011.  (unsupervised Neural Net) Real-time credit card fraud detection using computational intelligence. Quah. 2008	
11	Attribute Selection Discretizing Numeric Attributes. Principal Component Analysis (PCA)	7	Fraud Classification Using Principal Component Analysis of Rigits, Brockett, 2002.  A Survey of Outlier Detection Methodologies. Hodge & Austin, 2004.	TBA
12	Detecting Anomalies. Outlier Detection.	7	Unsupervised Outlier Detection in Time Series Data. Ferdousi, 2006.  Anomaly Detection: A Survey. Chandola. 2009.	TBA
13	Randomization vs. Boosting. Random Forest. Metalearners. Boosting. AdaBoost Alternating Decision Tree.	8	(random forests) Learned lessons in credit card fraud detection from a practitioner perspective. Pozzollo. 2014  Pattern Classification Using Support Vector Machine Ensemble. Kim et al. 2002.	TBA
14	Final Project Presentations			