

**IS 687**  
**Transaction Mining / Fraud Detection**  
**Course Syllabus**  
**Summer 2016**

**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 R, RStudio, and Weka data mining tools. 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, Ph.D.

**E-Mail:** [rudniy@njit.edu](mailto:rudniy@njit.edu)

**Skype:** alxrud

**Office Hours:** by appointment.

**Communication:** Email, Moodle forum, or Skype.

**Textbook:** Tan, Steinbach, Kumar. Introduction to Data Mining. 1st Edition, 2005, ISBN-13: 978-0321321367, ISBN-10: 0321321367.

This is a link to this book on Amazon: <http://www.amazon.com/Introduction-Data-Mining-Pang-Ning-Tan/dp/0321321367>

**Grading Policy:**

%	Assignment
10%	Moodle participation and quizzes
30%	Homework assignments
30%	Final project
30%	Final exam

**Topics and Reading Material**

Topic	Book Chapters	Papers for Reading and In-Class Presentation (will be posted on Moodle)	Homework
1 Electronic crime. Types of fraud. Introduction to data mining. Supervised vs. unsupervised learning. Industry standards. Benford's Law.	1, 2	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.  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.	Install R and RStudio, Install Weka
2 Concepts. Instances. Attributes. Preparing data. Sparse data, missing values. Weka Data Mining Workbench.	3	Data Mining for Fraud Detection: Toward an Improvement on Internal Control Systems? Jans, Lybaert, Vanhoof.  Automation and Disruption in Stolen Payment Card Markets. Peacock, Friedman. 13th Annual Workshop on	As posted on Moodle

			the Economics and Inf. Sec., Penn State U, 2014.	
3	Training, Testing, Cross-Validation. Lift Charts. ROC Curves. Precision-Recall curves. Cost-Sensitive Learning.	4.5, 4.6 from Ch.4; 5.7 from Ch. 5	ROC Graphs: Notes and Practical Considerations for Data Mining Researchers. Fawcett. 2003. Measuring classification performance: the hmeasure package. Anagnostopoulos, Hand, Adams, 2012.	As posted on Moodle
4	Linear models. Linear Regression, Logistic Regression.	Appendix D	(Regression) Data mining for credit card fraud: A comparative study. Bhattacharyya, Jha, Tharakunnel, Westland. 2011	As posted on Moodle
5	1R. Trees. Classification Rules. C4.5. Pruning. From trees to rules. Classification Rules. RIPPER.	4.1, 4.2, 4.3, 4.4 from Ch 4; 5.1 from Ch. 5	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. 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. A cost-sensitive decision tree approach for fraud detection. Sahin, Bulkan, Duman. Expert Systems with Applications 40 (2013) 5916–5923. (C4.5, RIPPER) Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection. Chan, Stolfo. 1998.	As posted on Moodle
6	K-nearest-neighbor. Naïve Bayes. Bayesian Classifiers. Bayesian Networks.	5.2, 5.3	(Bayes) Credit Card Fraud Detection: Personalized or Aggregated Model. Alowais, Soon. 2012.	As posted on Moodle
7	Neural Networks. Perceptrons. Backpropagation.	5.4	(backpropagation, bayes) Minority Report in Fraud Detection: Classification of Skewed Data. Phua, et al. 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 (Bayes net, neural net) Data Mining techniques for the detection of fraudulent financial statements. Kirkos et al. 2007.	As posted on Moodle
8	Support Vector Machines. Support Vector Regression.	5.5	(SVM) Detecting Click Fraud in Online Advertising: A Data Mining Approach. Oentaryo et al. 2013.	As posted on Moodle
9	Randomization vs. Boosting. Random Forest. Metalearners. Boosting. AdaBoost Alternating Decision Tree.	5.6	(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.	
10	Association Rules.	6	Association rules applied to credit card fraud detection. Sanchez, Vila, Cerda, Serrano. Expert Systems with Applications 36 (2009) 3630–3640.	As posted on Moodle

11	Cluster Analysis	8	(Clustering) Mining Information from Plastic Card Transaction Streams. Tasoulis, Adams, Weston, Hand. 18th Int. Conf. on Comp. Stat., Porto, Portugal, 2008.	As posted on Moodle
12	Attribute Selection Discretizing Numeric Attributes. Principal Component Analysis (PCA)	Appendix B	Fraud Classification Using Principal Component Analysis of Rredits, Brockett, 2002.	As posted on Moodle
13	Detecting Anomalies. Outlier Detection.	10	A Survey of Outlier Detection Methodologies. Hodge & Austin, 2004.  Unsupervised Outlier Detection in Time Series Data. Ferdousi, 2006.  Anomaly Detection: A Survey. Chandola. 2009.	As posted on Moodle
14	Final Exam			