

# Teaching Supervised and Unsupervised Learning Methods

Rajib L. Saha

Indian School of Business

# Courses I teach

- Business Analytics using Data Mining
  - Supervised and Unsupervised learning
  - Elective in full-time MBA program
- Data Mining – I
  - Unsupervised learning
  - Graduate-level, 15 months long, part-time program in analytics

# What I'm going to talk about today

1. Course curriculum in brief (to set the context)
2. Three course components in details
  - i. The framework for a course project
  - ii. Facilitate learning through peer evaluation – a framework for evaluation
  - iii. Offline videos to facilitate learning in a software-dependent course
3. Datasets I have been using for projects

# Course curriculum in brief

*(refer to uploaded document for details)*

## **Business Analytics using Data Mining (BADM)**

- **Supervised Learning**
  - Explaining vs. Predicting
  - Prediction and classification goals and performance evaluation
  - A bunch of prediction and classification methods
- **Unsupervised Learning**
  - Clustering
  - Principal Component Analysis
  - Association Rules

## **Data Mining – I (DMg-I)**

- **Unsupervised Learning**
  - K-Means and Hierarchical clustering
  - PCA and SVD as dimension reduction methods
  - Association Rules and Sequential Pattern Mining as methods for relationship mining
  - Recommender Systems
  - Network Analytics

# Predictive Analytics

- An emphasis on understanding ...
  - Explaining vs. Prediction goals
  - Choice of metrics for evaluating the predicting power of a method

# Prediction $\neq$ Explaining

- Choice of methods
- Choice of predictors
- The importance of cross validation
- Choice of evaluation metrics/evaluation of models

# I position a supervised learning method based on what it can do

Method	Numerical Y		Categorical Y			Numerical X	Categorical X	Lazy Learning	Variable Selection
	Explain	Predict	Profile	Classify	Ranking				
Linear Regression	Yes	Yes	No	No	No	Yes	Yes	No	Yes
Logistic Regression	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
k-NN	No	Yes	No	Yes	Yes	Yes	No	Yes	No
Naïve Bayes	No	No	No	Yes	Yes	No	Yes	No	No
Tree	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Ensembles	No	Yes	No	Yes	Yes	Depends	Depends	Depends	No

# Evaluation of classification models

## **Goal**

- Profiling (explaining) vs. Classification (prediction)
- Classifying vs. ranking

## **Adding Information**

- Asymmetric class importance
- Asymmetric misclassification cost

## **Classification performance**

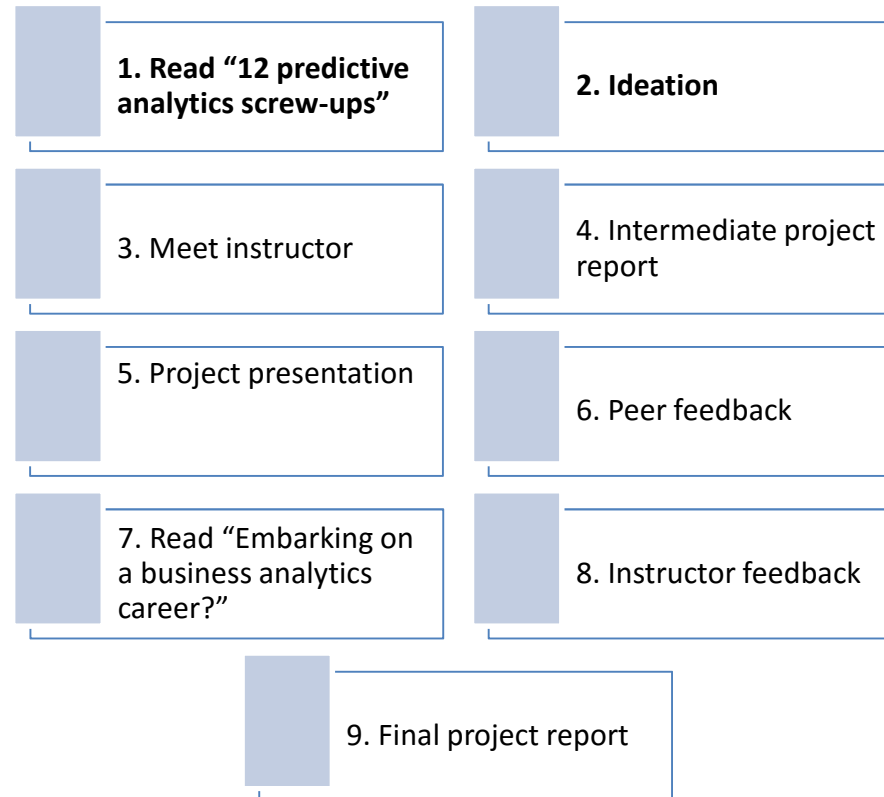
- Confusion matrix
- Sensitivity, Specificity, Precision, F1-Score, and ROC curve
- Lift Chart (for ranking)



# Project Framework Predictive Analytics

- Project ideation
- Intermediate project report
- Project presentation
- Final project report

# Execution of the project



# 12 predictive analytics screw-ups

- *Refer to the uploaded article*

# Project Ideation

## 1/5. Business Goal

- Who is the stakeholder or the client with regard to the business objective you are proposing?
- A description of the business objective? What are the benefits of implementing this idea? What opportunity is it creating? What shortcoming does it address?
- What sorts of decisions (by the stakeholder) are involved in achieving this goal?
- What would be considered a success? How would you quantify success?

# Project Ideation

## 2/5. Analytics/Data Mining goal

- Provide a description of the analytics objective/ data mining problem.
- Is this a supervised or unsupervised task? Is it predictive or descriptive? Is it retrospective or forward-looking?
- What is the main outcome variable(s) of interest?
- What predictors are needed?

# Project Ideation

## 3/5. Data Preparation

- Provide a brief description of the available data.
- Provide some guidance on the dataset that will be used and the pre-processing or data preparation that might be needed before you can build a model to address the data mining problem.
- Provide a sample of ten rows (records) with column (variable) names that will be used --- that is, how the dataset is going to look post data preparation phase and before you build your models.

# Project Ideation

## 4/5. Methods and Evaluation

- What are some data mining methods to consider?
- Which performance measures are appropriate? How do they align with the business goal?

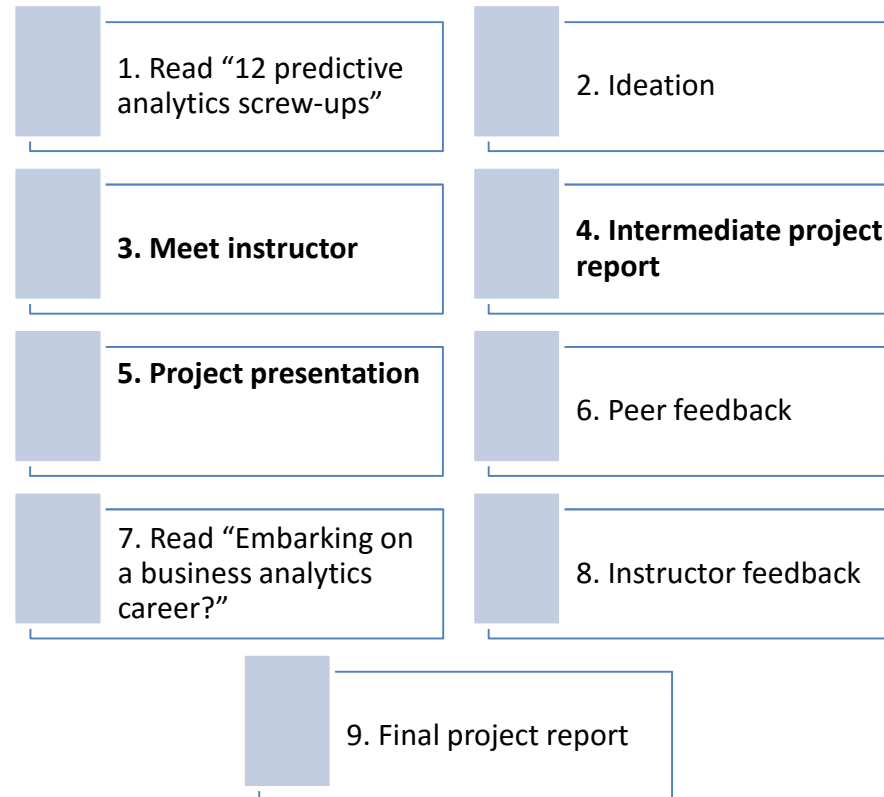
# Project Ideation

## 5/5. Implementation/Recommendation

- Operational requirements or constraints (for example, will the solution be run in real-time? will it require collecting new data? will it be a one-time analysis or ongoing?)



# Execution of the project



# Peer evaluation

A framework for evaluating projects

- Project on predictive analytics

*Refer to uploaded document for details*

# Embarking on a Business Analytics Career?

- *Refer to the uploaded article*

# Project Framework

## Unsupervised Learning

- Describe the key **insight** in not more than one sentence.
- **Steps** (data processing, tools used, parameters chosen, etc.) to reproduce the insight
- **Social and/or business values** of the insight(s) to specific **stakeholders**

# Peer evaluation

A framework for evaluating projects

- A project on using unsupervised learning methods

*Refer to uploaded document for details*

# Offline video content for software-dependent course

- *Refer to a sample video*
  - *Running\_classification\_tree.mp4*



TechSmith Camtasia™ 9

# Public datasets I've used for projects and assignments, so far

- Data from hosted competitions on Kaggle.com
- UCI machine learning repository
- Crime data from Chicago data portal [data.cityofchicago.org/](http://data.cityofchicago.org/)
- International trade-flow data from <http://www.wto.org/>

Thank you