[INDIANA UNIVERSITY BLOOMINGTON](http://www.indiana.edu/)

[](http://www.indiana.edu/)

[**CSCI-B565 Data Mining**](https://www.cs.indiana.edu/~predrag/classes/2018springb565/)

* [Weekly Schedule](https://www.cs.indiana.edu/~predrag/classes/2018springb565/)

* [Syllabus](https://www.cs.indiana.edu/~predrag/classes/2018springb565/syllabus.html)

* [Contact Us](https://www.cs.indiana.edu/~predrag/classes/2018springb565/contact.html)

**Announcements**

* ~~• No classes: 03/13, 03/15 (Spring break)~~
* ~~• Midterm exam: March 8, in class.~~
* • Final exam: May 3, 12:30-2:30pm. IU schedule available [here](http://enrollmentbulletin.indiana.edu/pdf/spring-exam-schedule.pdf).
* • Project report is due 4/28. Instructions: [here](https://www.cs.indiana.edu/~predrag/classes/2018springb565/hw6.pdf)
* • Office hours after the final exam: May 4, 6-8pm ☀️

**———————————————————————**

Last updated: April 29, 2018

**Weekly Schedule**

**———————————————————————**

**Week 16, April 23**

**Topics**

* • Privacy-preserving data mining
* • Review for final exam

**Reading materials**

* • A book chapter by Stan Matwin [here](https://link.springer.com/chapter/10.1007/978-3-642-30487-3_11)
* • ICML 2010 slides by Stan Matwin [here](http://www.site.uottawa.ca/~stan/icml10tutPrivacy/ICML2010-privacy.pdf)
* • Modified slides used in class [here](https://www.cs.indiana.edu/~predrag/classes/2018springb565/ppdm.pdf)

**———————————————————————**

**Week 15, April 16**

**Topics**

* • Support vector machines

**Reading materials**

* • Textbook #1: Classification: alternative techniques (Chapter 5)
* • Textbook #2: SVM: support vector machines (Chapter 3)

**Handouts and code**

* • Slides available [here](https://www.cs.indiana.edu/~predrag/classes/2018springb565/svm.pdf)

**Homework assignment #5:**[here](https://www.cs.indiana.edu/~predrag/classes/2018springb565/hw5.pdf).

**———————————————————————**

**Week 14, April 9**

**Topics**

* • Introduction to support vector machines
* • Committee machines

**Reading materials**

* • Textbook #1: Classification: alternative techniques (Chapter 5)

**———————————————————————**

**Week 13, April 2**

**Topics**

* • Rule-based classifiers
* • Instance-based classifiers
* • Introduction to linear classifiers and neural networks

**Reading materials**

* • Textbook #1: Classification: alternative techniques (Chapter 5)
* • Textbook #2: kNN: k-Nearest neighbors (Chapter 8)
* • Textbook #2: Naive Bayes (Chapter 9)

**Handouts and code**

* • Chapter 5 [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/ch5.pdf)

**———————————————————————**

**Week 12, March 26**

**Topics**

* • Mining sequences

**Reading materials**

* • Textbook #1: Association analysis: advanced concepts (Chapter 7)

**Handouts and code**

* • Advanced association rule mining [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/ch7.pdf)
* • Graphlet counting [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/graphlets.pdf)

**———————————————————————**

**Week 11, March 19**

**Topics**

* • Association rule mining

**Reading materials**

* • Textbook #1: Association analysis (Chapter 6)
* • Textbook #2: Apriori (Chapter 4)

**Homework assignment #4:**[here](https://www.cs.indiana.edu/~predrag/classes/2018springb565/hw4.pdf).

**———————————————————————**

**Week 10, March 12**

**Spring break.**

**———————————————————————**

**Week 9, March 5**

**Topics**

* • Association rule mining
* • Midterm exam on Thursday

**Handouts and code**

* • Association rule mining [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/ch6.pdf)

**Reading materials**

* • Textbook #1: Association analysis (Chapter 6)
* • Textbook #2: Apriori (Chapter 4)

**———————————————————————**

**Week 8, February 26**

**Topics**

* • The expectation maximization (EM) algorithm
* • Class prior estimation

**Handouts and code**

* • EM slides [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/parameter_estimation_full.pdf) (contains slides from previous week)
* • Class prior presentation [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/shantanu.pdf)

**Reading materials**

* • Textbook #1:Cluster analysis: basic concepts and algorithms (Chapter 4)
* • Textbook #2: EM (Chapter 5)

**———————————————————————**

**Week 7, February 19**

**Topics**

* • Parameter estimation

**Handouts and code**

* • Parameter estimation [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/parameter_estimation.pdf)

**Reading materials**

* • Textbook #1: Data (Chapter 2)

**Homework assignment #3:**[here](https://www.cs.indiana.edu/~predrag/classes/2018springb565/hw3.pdf)

**———————————————————————**

**Week 6, February 12**

**Topics**

* • Evaluating clustering
* • Data pre-processing issues

**Handouts and code**

* • Data pre-processing [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/ch2.pdf)

**Reading materials**

* • Textbook #1: Data (Chapter 2)
* • Web site from Stanford NLP [here](https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-clustering-1.html)

**———————————————————————**

**Week 5, February 5**

**Topics**

* • K-means clustering, accelerations
* • Hierarchical clustering

**Handouts and code**

* • K-means acceleration algorithm [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/kmeansaccelerated.pdf)

**Reading materials**

* • K-means acceleration papers: [Elkan 2003](https://www.cs.indiana.edu/~predrag/classes/2018springb565/papers/elkan_icml_2003.pdf), [Arthur 2006](https://www.cs.indiana.edu/~predrag/classes/2018springb565/papers/arthur_soda_2006.pdf), [Hamerly 2010](https://www.cs.indiana.edu/~predrag/classes/2018springb565/papers/hamerly_sdm_2010.pdf), [Hamerly 2014](https://www.cs.indiana.edu/~predrag/classes/2018springb565/papers/hamerly_bookchapter_2014.pdf), [Bachem 2016](https://www.cs.indiana.edu/~predrag/classes/2018springb565/papers/bachem_nips_2016.pdf)

**Homework assignment #2:** [here](https://www.cs.indiana.edu/~predrag/classes/2018springb565/hw2.pdf)

**———————————————————————**

**Week 4, January 29**

**Topics**

* • K-means clustering

**Reading materials**

* • Textbook #1: Cluster analysis: basic concepts and algorithms (Chapter 4)

**Handouts and code**

* • Chapter 8 [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/ch8.pdf)

**———————————————————————**

**Week 3, January 22**

**Topics**

* • Evaluation of classification models
* • Tom Dietterich’s slides, modified (with permission) by Sriraam Natarajan [here](https://www.cs.indiana.edu/~predrag/classes/2018springb565/eval.pdf)

**Reading materials**

* • Textbook #1: Basic classification (Chapter 4)

**———————————————————————**

**Week 2, January 15**

**Topics**

* • Decision trees

**Reading materials**

* • Textbook #1: Basic classification (Chapter 4)
* • Textbook #2: C4.5 (Chapter 1)

**Handouts and code**

* • Chapter 4 [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/ch4.pdf)

**Homework assignment #1:** [here](https://www.cs.indiana.edu/~predrag/classes/2018springb565/hw1.pdf)

**———————————————————————**

**Week 1, January 8**

**Topics**

* • Class overview and logistics
* • Introduction to data mining

**Reading materials**

* • Textbook #1: Introduction (Chapter 1)

**Handouts and code**

* • Class overview [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/intro.pdf)
* **•**Chapter 1 [slides](https://www.cs.indiana.edu/~predrag/classes/2018springb565/ch1.pdf)

# [INDIANA UNIVERSITY BLOOMINGTON](http://www.indiana.edu/)

[](http://www.indiana.edu/)

# [CSCI-B565 Data Mining](https://www.cs.indiana.edu/~predrag/classes/2018springb565/)

* [Weekly Schedule](https://www.cs.indiana.edu/~predrag/classes/2018springb565/)

* [Syllabus](https://www.cs.indiana.edu/~predrag/classes/2018springb565/syllabus.html)

* [Contact Us](https://www.cs.indiana.edu/~predrag/classes/2018springb565/contact.html)

**Class Meets**

When: Tuesdays and Thursdays, 11:15am-12:30pm

Where: Fine Arts 102

**Course Objectives**

Data mining is a practical discipline that aims to identify interesting patterns and relationships hidden in data. It emerged as one of the most dynamic fields of computer science as the new relationships discovered in the data can be used to predict future. This course is designed to introduce fundamental concepts of data mining and provide hands-on experience with several techniques. The students will be expected to develop a broad background in the field of data mining and develop skills to solve practical problems. The problems will be presented from various fields, such as fraud detection, e-commerce, stock market, medicine and life sciences.

**Class materials**

**Textbook #1:** Introduction to Data Mining by Tan et al., Pearson 2006.

**Textbook #2:** The Top Ten Algorithms in Data Mining by Wu and Kumar, CRC Press 2009.

**Recommended book:** Data Mining: Concepts and Techniques by Han et al., Morgan Kaufmann 2006.

**Supplementary materials:** to be provided in class.

**Topics**

* • introduction to data mining
* • data representation and data preprocessing
* • data visualization
* • mining association rules
* • classification and regression methods
* • model selection and evaluation
* • clustering
* • social issues in data mining; privacy-preserving  data mining
* • case studies on various types of data

**Grading**

Midterm exam: 20%

Final exam: 20%

Homework assignments: 35%

Mini project: 20%

Class participation: 5%

**Late Policy and Academic Honesty**

All assignments and exams are individual, except when collaboration is explicitly allowed. All the sources used for problem solution must be acknowledged, e.g. web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously; for detailed information see [Indiana University Code of Student Rights, Responsibilities, and Conduct](http://studentcode.iu.edu/).

Last updated: January 14, 2018

# [INDIANA UNIVERSITY BLOOMINGTON](http://www.indiana.edu/)

[](http://www.indiana.edu/)

# [CSCI-B565 Data Mining](https://www.cs.indiana.edu/~predrag/classes/2018springb565/)

[Weekly Schedule](https://www.cs.indiana.edu/~predrag/classes/2018springb565/)

[Syllabus](https://www.cs.indiana.edu/~predrag/classes/2018springb565/syllabus.html)

[Contact Us](https://www.cs.indiana.edu/~predrag/classes/2018springb565/contact.html)

**Instructor:**

# [Predrag Radivojac](http://www.cs.indiana.edu/~predrag)

**Office:** Luddy Hall 2048

**Email:** predrag@indiana.edu

**Phone:** 812-856-1851

**Office Hours:** Tue/Thu 2-3pm

**Associate Instructor(s):**

Moses Stamboulian

Email: mstambou

Office Hours:

MW: 4-5:30pm in Luddy Hall 2052

﻿​

Eriya Terada

Email: eterada

Office Hours:

TR: 9:30-11am in Luddy Hall 1069

﻿​

Benjamin Rosenzweig

Email: bkrosenz

Office Hours:

TR: 3-4:30pm in Luddy Hall 1069

﻿​

Yuxiang Jiang

Email: yuxjiang

Office Hours:

MW: 9:30-11am in Luddy Hall 2069

﻿​

**\*Note:** due to the slow process in IU room reservation system, changes will be posted here during the semester. Please check this web page regularly.

https://maps.gstatic.com/mapfiles/api-3/images/spotlight-poi2.png



