

# Doing Data Science Using Machine Learning Models

A short course for First National Bank, Omaha, NE

Spring 2018

**Instructor:** Dr. Mahbubul Majumder      Office: DSC 238  
Phone: (402) 554-2734      Email: mmajumder@unomaha.edu

**Lecture time:** Wednesday 2:30PM - 4:30PM in Mammel Hall (room 115)

**Office Hours:** After class or by appointment

**Course web page:** <http://mamajumder.github.io/first-national>

**Text books:** Notes and various resources will be provided during lectures.

**Recommended books:**

1. Modern Data Science with R by Baumer et. al.
2. The Elements of Statistical Learning by Hastie et. al.
3. An Introduction to Statistical Learning with Applications in R by James et. al.

## Course outline (approximate)

Session	Chapter Description	Date
1	Introduction to data science and reproducible report using R	Jan 24, 2018
2	Effective display of data and creating interactive visualization	Feb 14, 2018
<b>Abstract of the project due: Mar 1, 2018</b>		
3	Gathering and manipulation of structured and date data	Mar 7, 2018
4	Gathering and manipulation of unstructured and text data	Mar 28, 2018
5	Overview of Statistical vs machine learning models	April 18, 2018
6	Assessment, selection and inference of machine learning models	May 9, 2018
<b>First draft of the project due: May 20, 2018</b>		
7	Models for supervised learning	May 30, 2018
8	Models for unsupervised learning	Jun 20, 2018
9	Model averages and ensemble modeling	Jul 11, 2018
10	Big data tools: hadoop, mapreduce, pig	Aug 8, 2018
<b>Final project due: Aug 8, 2018</b>		
<b>Project presentation at Bank Head Quarter: Date and time TBD</b>		

**Practical Assignments:** There will be five practical data assignments throughout the course. The main purpose of these assignment questions is to challenge participants with micro level data science problems. Understanding those problems should help to efficiently develop a data analysis report in a timely manner.