Syllabus
EGGG 367: Data Science 1
FALL 2019

Instructor: Dr. Nii Attoh-Okine, 354 Dupont Hall/109 Evans Hall, email: okine@udel.edu
Time and Location: Th 3:30 pm - 6:30 pm [Dupont 350]
                     Tuesday       10:30am - 11:30 am
                     Wednesday    12:00 Noon - 1:30pm
                     Thursday     1:30pm - 2:30 pm
Office Hours:       Wednesday 12:00 Noon - 1:30pm
                     Thursday 1:30pm - 2:30 pm
Grading: There will be weekly assignments and project. They will count toward the grade
as follows.

   Assignments  60%
   Midterm       15%
   Final         25%

Description of the Course

Advances in technology have allowed us to collect massive amounts of data. A data
scientist is a person who has the skills, knowledge, and ability to extract actionable knowledge
from the data either for good of society, advancement of science and technology, promotes
business, etc. The course will examine the central question of “what is Big Data” and the
how can engineers, statisticians, health related specialists, computer scientists, social scientist
and other professionals can employ tools and techniques of data science. This course will
cover the topics needed to solve data science problems, which include data preparation, data
characterization and presentation, data analysis, and data products. Enrolling in this course
will help the student develop a deeper understanding of the various phases of Big Data.

Syllabus

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Overview and Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Getting Started with R (R-studio)</td>
</tr>
<tr>
<td>3</td>
<td>Exploratory Data Analysis (ggplot2)</td>
</tr>
<tr>
<td>4</td>
<td>Data Visualization (Univariate, Bivariate, Multivariate)</td>
</tr>
<tr>
<td>5</td>
<td>Clustering (Distance Measures, Hierarchical, Partitioned-Based Clustering)</td>
</tr>
<tr>
<td>6</td>
<td>Predictive Analytics (Predictive Modeling, Multiple Linear Regression, ROC curves)</td>
</tr>
<tr>
<td>7</td>
<td>Predictive Analytics (Logistic Regression, Naive Classifiers)</td>
</tr>
<tr>
<td>8</td>
<td>Deep Learning</td>
</tr>
<tr>
<td>9</td>
<td>Applications (Case Studies), Jupyter</td>
</tr>
</tbody>
</table>