Instructor: Jacob Turner, Ph.D.
Department: Mathematics and Statistics
Email: turnerja2@sfasu.edu
Phone: 936-468-1692
Office: 342 NM
Office Hours:

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>10am-11:30am</td>
<td></td>
<td></td>
<td></td>
<td>10am-11:30am</td>
</tr>
</tbody>
</table>

You can also meet with me by appointment.

Class meeting time and place: TBD

Text and Materials:

<table>
<thead>
<tr>
<th>Introduction to Statistical Learning</th>
<th>James, Witten, Hastie, &amp; Tibshirani</th>
<th>Free Online</th>
<th>Springer</th>
</tr>
</thead>
<tbody>
<tr>
<td>R statistical software</td>
<td></td>
<td>Free Download Online</td>
<td></td>
</tr>
</tbody>
</table>

Course Details:

**Course Objectives:**
A broad overview of various statistical methods with an emphasis in predictive modeling and R programming. Topics will include statistical learning concepts and terminology, classification, advanced feature selection methods, bootstrap and cross validation procedures, tree algorithms, unsupervised learning. In addition to the general skill set, a great emphasis will be placed on understanding how to conduct an analysis workflow from start to finish, from exploratory work, model development, and validation.

**Course outline (Approximate time spent)**

**Course Outline:**

- What is statistical learning?
  - Exploratory Analysis
  - Key Definitions
  - Need for Validation
  - The Bias/Variance Trade Off

- Multiple Linear Regression
  - Estimation and assumptions
  - Pitfalls and Short comings
  - K-Nearest Neighbors

Approximate time spent: 15%
• Classification 25%
  o Logistic Regression
  o Linear Discriminate Analysis
  o Intro to ROC Curves
• Resampling Methods 15%
  o Cross Validation
  o Bootstrap
  o Feature Selection Using LASSO
• Tree Based Methods 15%
  o CART models
  o CART vs Regression
  o Random Forrest
• Unsupervised Learning and Data Visualization 15%
  o Clustering methods
  o Intro to PCA
  o Visualizations for Unsupervised learning

Student Learning Outcomes (SLO): At the end of STAT 328, a student who has studied and learned the material should be able to:
1. Exhibit a firm understanding of how to enter, manipulate, analyze, and visualize data within R statistical software.
2. Identify and apply correct modeling procedures and strategies to different data types depending on the study design
3. Assess model and prediction performance through resampling techniques and correct interpretation of ROC curves
4. Explain the general concept of the trade offs between model complexity and interpretability
5. Explain the difference between how resampling can play a role in feature selection and how that potentially differs from estimating prediction performance
6. Apply appropriate unsupervised analysis techniques to aid in data descriptions, reduction, and visualization
7. Use machine learning algorithms such as randomForest or support vector machines as an alternative to traditional predictive modeling techniques.

Grading: The final average will be computed using the following weights:
- Homework and additional to dos 40%
- 1-2 projects (Group contest) 40%
- Midterm/Final 20%

Homework
Homework assignments will be given out of the text and potentially some standalone work that I will provide. It is very important that you stay up on trying to master some of your R capabilities and the HW will help you to practice. I highly recommend you take a look at the end of each chapter in the text book. The authors provide very nice R examples and tutorials for many of the things that we will be doing.
Midterm and Final:
The midterm and final will focus on definitions, key statistical learning concepts, small mathematical derivations, and the general information about each method covered. Depending on how the class go, some material may be based on a few assigned readings. For this independent study, these will be “take” home with an appropriate amount of time allocated.

Projects:
There will be two projects during the course. The first project will be more technical in nature to help facilitate the early concepts and techniques of the course. The second project will be a contest among the groups on a big data modeling and prediction. Dr. Turner will have a hold out data set to assess each groups predictive capabilities. Bonus points will be allotted for the winning group! For each project, the students will present their work and approach to the class via a ppt presentation. R code must be provided as a secondary document.

Grading Scale:

<table>
<thead>
<tr>
<th>Letter Grade</th>
<th>% Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90-100</td>
</tr>
<tr>
<td>B</td>
<td>80-89</td>
</tr>
<tr>
<td>C</td>
<td>70-79</td>
</tr>
<tr>
<td>D</td>
<td>60-69</td>
</tr>
<tr>
<td>F</td>
<td>below 59</td>
</tr>
</tbody>
</table>

Expectations and Class Rules:
ALL cell phones are to be turned OFF and put away. If someone violates this rule (texting, playing games, etc.), then the person must leave the class room for that class period without any further delay according to the professor’s instructions. No food in the classroom. Don’t leave the classroom in the middle of the lecture.

Attendance Policy:
It is important in an independent study to remain engaged with the course work and regularly have conversations with me. You talking to me throughout the course is expected.

Formalities Which Must Be Included in This Document:

Number of credit hours for this course 3
The following is an excerpt from SFA Policy 5.4:

The federal definition of a credit hour is an amount of work represented in intended learning outcomes and verified by evidence of student achievement that is an institutionally established equivalency that reasonably approximates:

1. Not less than one hour of classroom or direct faculty instruction and a minimum of two hours out-of-class student work each week for approximately fifteen weeks for one semester or trimester hour of credit, or 10 to 12 weeks for one quarter hour of credit, or the equivalent amount of work over a different amount of time, or;

2. At least an equivalent amount of work as outlined in item 1 above for other academic activities as established by the institution including laboratory work, internships, practica, studio work, and other academic work leading to the award of credit hours.

To this end, all students in courses offered by the Department of Mathematics and Statistics that wish to be successful should plan to spend a minimum of two hours outside of class for every credit hour associated with this course. Expected activities to be completed in the time outside of class include reviewing notes from previous class meetings, reading assigned course resources, completing all assigned exercises and projects, and performing periodic assessment preparation.
Academic Integrity (A-9.1)

The penalty for violating Academic Integrity policy A-9.1 at any time during this semester is failure of the course. No exceptions. No grade will be calculated for a student who violates the policy. They will be asked to sign an academic dishonesty form after evidence of their violation has been provided to them and they will receive an F in the course no matter what grades have been accumulated to that point in the semester.

Academic integrity is a responsibility of all university faculty and students. Faculty members promote academic integrity in multiple ways including instruction on the components of academic honesty, as well as abiding by university policy on penalties for cheating and plagiarism.

Definition of Academic Dishonesty
Academic dishonesty includes both cheating and plagiarism. Cheating includes but is not limited to (1) using or attempting to use unauthorized materials to aid in achieving a better grade on a component of a class; (2) the falsification or invention of any information, including citations, on an assigned exercise; and/or (3) helping or attempting to help another in an act of cheating or plagiarism. Plagiarism is presenting the words or ideas of another person as if they were your own. Examples of plagiarism are (1) submitting an assignment as if it were one's own work when, in fact, it is at least partly the work of another; (2) submitting a work that has been purchased or otherwise obtained from an Internet source or another source; and (3) incorporating the words or ideas of an author into one's paper without giving the author due credit.

Please read the complete policy at http://www.sfasu.edu/policies/academic_integrity.asp

Withheld Grades Semester Grades Policy (A-54)

Ordinarily, at the discretion of the instructor of record and with the approval of the academic chair/director, a grade of WH will be assigned only if the student cannot complete the course work because of unavoidable circumstances. Students must complete the work within one calendar year from the end of the semester in which they receive a WH, or the grade automatically becomes an F. If students register for the same course in future terms the WH will automatically become an F and will be counted as a repeated course for the purpose of computing the grade point average.

The circumstances precipitating the request must have occurred after the last day in which a student could withdraw from a course. Students requesting a WH must be passing the course with a minimum projected grade of C.

Students with Disabilities

To obtain disability related accommodations, alternate formats and/or auxiliary aids, students with disabilities must contact the Office of Disability Services (ODS), Human Services Building, and Room 325, 468-3004 / 468-1004 (TDD) as early as possible in the semester. Once verified, ODS will notify the course instructor and outline the accommodation and/or auxiliary aids to be provided. Failure to request services in a timely manner may delay your accommodations. For additional information, go to http://www.sfasu.edu/disabilityservices/


**Acceptable Student Behavior**

Classroom behavior should not interfere with the instructor’s ability to conduct the class or the ability of other students to learn from the instructional program (see the Student Conduct Code, policy 10.4). Unacceptable or disruptive behavior will not be tolerated. Students who disrupt the learning environment may be asked to leave class and may be subject to judicial, academic or other penalties. This prohibition applies to all instructional forums, including electronic, classroom, labs, discussion groups, fieldtrips, etc. The instructor shall have full discretion over what behavior is appropriate/inappropriate in the classroom. Students who do not attend class regularly or who perform poorly on class projects/exams may be referred to the Early Alert Program. This program provides students with recommendations for resources or other assistance that is available to help SFA students succeed.