Remote / Online Course Syllabus
GISC6323: Machine Learning for Socio-Economic and Geo-Referenced Data
Spring 2021

Professor Contact Information

Professor	Michael Tiefelsdorf
Email Address	tiefelsdorf@utd.edu
Online Office Hours	By appointment or when online on MS Teams (subject to availability)

Course Modality and Expectations

Instructional Mode	Online on Thursdays 1:00-3:45 pm and remote after the lectures in the General tab of MS Teams.
Course Platform	MS Teams is used for course delivery and as lecture notes depository whereas eLearning is used for lab assignments and final exam.
Expectations	Interactive online participation is preferred and will enhance the learning experience.

COVID-19 Guidelines and Resources

The information contained in the following link lists the University's COVID-19 resources for students and instructors of record.

Please see http://go.utdallas.edu/syllabus-policies.

Class Recordings

Students are expected to follow appropriate University policies and maintain the security of passwords used to access recorded lectures. Unless the Office of Student AccessAbility has approved the student to record the instruction, students are expressly prohibited from recording any part of this course. Recordings may not be published, reproduced, or shared with those not in the class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. Failure to comply with these University requirements is a violation of the Student Code of Conduct.

The instructor may record meetings of this course. Any recordings will be available to all students registered for this class as they are intended to supplement the classroom experience. Students are expected to follow appropriate University policies and maintain the security of passwords used to access recorded lectures. Unless the Office of Student AccessAbility has approved the student to record the instruction, students are expressly prohibited from recording any part of this course. Recordings may not be published, reproduced, or shared with those not in the class, or uploaded to other online environments except to implement an approved Office of Student AccessAbility accommodation. If the instructor or a UTD school/department/office plans any other uses for the recordings, consent of the students identifiable in the recordings is required prior to such use unless an exception is allowed by law. Failure to comply with these University requirements is a violation of the Student Code of Conduct.

Course Pre-requisites, Co-requisites, and/or Other Restrictions

GISC6301 or equivalent statistical analysis and modeling skills including linear regression.

A general understanding of computer use is expected. In particular, working knowledge of the course software is expected.

Required Texts

[**Gar**] Gareth, James, Daniela Witten, Trevor Hastie and Robert Tibshirani, 2013. *An Introduction to Statistical Learning with Applications in R*. Springer Verlag

Notes: Available online at https://statlearning.com/. A comprehensive and thorough introduction to most machine learning concepts and techniques with a rich set of examples. Its implementations are a little outdated. However, neural networks and deep learning are not included.

[**Boe**] Boehmke, Bradley and Brandon Greenwell, 2020. *Hands-On Machine Learning with R*. CRC Press.

Notes: Available online at https://bradleyboehmke.github.io/HOML/. Many examples of machine learning techniques using . It follows the more modern paradigm of data sciences.

Supplemental Texts

[**Wic**] Wickham, Hadley and Garrett Grolemund, 2017. *R for Data Science. Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly

Notes: Available online at https://r4ds.had.co.nz . Introduction to from a data science perspective. Selected chapters will be used.

[**Col**] Collet, Francois and J.J. Allaire. 2018. *Deep Learning with R*. Manning

Notes: Code examples are available at https://github.com//jjallaire/deep-learning-with-r-notebooks . Makes us of the professional keras and tensorflow libraries and GPUs.

Course Description

Models and algorithms as well as their underlying conceptional foundations to structure and predict socio-economic and geo-referenced data are introduced. Model refinements and limitations are discussed. Open-source software and commonly available hardware are used. Practical examples of [a] supervised machine learning to develop classification rules as well as to perform predictions and [b] unsupervised data mining tools to uncover a hidden organization of data objects are used to explore the strength and weaknesses of selected data analytical methods and to examine the resulting output.

Student Learning Objectives/Outcomes

Upon completing this class, students:

- *have a working understanding of the underlying concepts of probabilistic, kernel, decision tree, support vector machines and neural network machine learning algorithms,*
- *have a working understanding of clustering algorithms,*
- *perform independently machine learning tasks,*
- *will be able to evaluate performances as well as risks of under- and overfitting of machine learning solutions,*
- *understand software and hardware developments in machine learning,*
- *being prepared to branch into advanced machine learning tasks such as deep learning.*

Academic Calendar

Date	Topic	Reading	Assignment Hand-outs
Jan 21	Introduction		
Jan 28	Data science tools in R	Wic05, Wic12 & Wic18	
Feb 04	Machine learning overview and kNN	Gar01 & Gar02	
Feb 11	Modeling Process & Feature and Target Engineering	Boe02 & Boe03	Lab01
Feb 18	Linear regression and kNN	Gar03, Boe04 & Boe08	
Feb 25	Classification	Gar04 & Boe05	Lab02
Mar 04	Unsupervised learning I: K-Means and principal components	Gar10, Boe17 & Boe20	
Mar 11	Unsupervised learning II: Hierarchical cluster analysis with and without constraints and model-based clustering	Gar10, Boe21 & Boe22	Lab03
Mar 18	**Spring Break**		
Mar 25	Resampling Methods	Gar05	
Apr 01	Linear Model Selection and Regularization	Gar06	
Apr 08	Moving Beyond Linearity. Regression splines (?)	Gar07 & Boe07	Lab04
Apr 15	Tree-Based Methods	Gar08, Boe09, Boe10 & Boe11	
Apr 22	Support Vector Machines	Gar09 & Boe14	Lab05
Apr 29	Neural Networks and Autoencoders	Boe13 & Boe19	
May 06	Review		
TBA	**Final Exam**		

Course Policies

Grading (credit) Criteria

Policies:

- Labs and the final exam need to be solved ***individually***. Plagiarism ***cannot*** be tolerated.
- Participation is highly encouraged but will not be graded. Engagement with the course material will lead to participation and indirectly to an increased comprehension of the course material.

Requirements	*Points*
5 Labs *@ 13 pts: labs should be handed online to the Labs folders in eLearning. The labs will prepare you in parts for the final exam.*	65 pts

	Final Exam: 35 pts *Open book and open notes.*
Late Work	Work that is late by *one day* will lead to a deduction of 10% of its points. Work that is late by *two days* will lead to a deduction of 20 % of its points. Later work *cannot be accepted* unless special circumstances can be claimed. Preferably contact the instructor before the deadline if you think that you may need to hand your assignment in late.
UTD Syllabus Policies	All UTD syllabus policies apply to this course. It is advisable to study these policies at least once per academic year. See http://go.utdallas.edu/syllabus-policies for details

Comet Creed

This creed was voted on by the UT Dallas student body in 2014. It is a standard that Comets choose to live by and encourage others to do the same:

"As a Comet, I pledge honesty, integrity, and service in all that I do."

The descriptions and timelines contained in this syllabus are subject to change at the discretion of the Professor.