Course	**GISC6323: Machine Learning for Socio-Economic and Georeferenced Data**
Professor	Michael Tiefelsdorf, Ph.D.
Term	Fall Semester 2019
Lectures & Labs	Monday 4:00 pm – 5:15 pm in GR3.602 Wednesday 4:00 pm – 5:15 pm in GR3.602

Contact Information

Office Phone	(972) 883-4954 (email correspondence is highly preferred)
Office Location	GR3.204
Email Address	tiefelsdorf@utdallas.edu Please start the **subject line** of each email with **ML** to alert the instructor that it is a GISC6301 related email and not junk mail.
Office Hours	Thursdays 3:00-5:00 pm in my office at GR3.204 and by appointment or for short drop-ins when my office door is open.
Other Information	Please check the ELEARNING's course site for lecture notes, quizzes, announcements, the discussion forum etc. on a regular basis.

Teaching Assistant	Yalin Yang
Office Location	GR3.414
Email Address	Yalin.Yang@utdallas.edu Please start the **subject line** of each email with **ML**
Office Hours	Mondays and Wednesdays 3:00-4:00 pm and by appointment

General Course Information

Pre-requisites, Co-requisites, & other restrictions

A prior undergraduate course in basic statistical analysis is highly recommended (e.g., SOCS 3405) as well as knowledge about the open source system 
Ability to operate a personal computer, data handling skills, ability to use the University Library, **ELEARNING** and internet resources is required. No prior GISciences knowledge is necessary.

Course Description

Models and algorithms as well as their underlying conceptional foundations to structure dynamic socio-economic and geo-referenced data are introduced. Open-source software and commonly available hardware are used. Practical examples of [a] supervised machine learning to develop classification rules and [b] unsupervised data mining 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. Where appropriate, ethical ramifications are discussed.

Learning Outcomes

Upon completing this class, students:

- *have a working idea of the underlying concepts of probabilistic, kernel, decision tree and neural network machine learning 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, and*
- *be aware of societal benefits and implications of machine learning.*

Required Texts & Materials

Gareth, James, Daniela Witten, Trevor Hastie and Robert Tibshirani, 2017. *An Introduction to Statistical Learning with Applications in R*. 8th corrected printing. Springer Verlag
Notes: Available online at UTD's library. A comprehensive and thorough introduction to most machine learning concepts and techniques with a rich set of examples. However, neural networks and deep learning are not included.

.

Supplemental Texts, Readings & Materials

Burkov, Andriy, 2019. *The Hundred-Page Machine Learning Book*. Self-published.
Notes: Available complementary online at the author's website and at Amazon for ~$ 36. The underlying ideas of all relevant techniques are succinctly introduced. Code examples are not given but references are made to Phyton's SCIKIT-learn.

Cathy O'Neil, 2016. *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group
Notes: Available online at UTD's library. This text reflects judiciously on big data analysis and machine learning outcomes for society. Course participants will critical review and summarize issues raised in individual chapters.

Brett Lantz, 2019. *Machine Learning with R. Expert techniques for predictive modeling.* 3rd edition. PackIt Publishing Ltd.
Notes: Available at Amazon for ~ $ 36. A light introduction to the use of machine learning algorithms in R with code examples. Avoids underlying mathematical and statistical foundation of techniques. Last chapter provides an overview of relevant computational issues for big data.

Chollet, Francois and J.J. Allaire, 2018. *Deep Learning with R*. Manning Publications Co.
Notes: The book is offered is several formats including online for a fee and abridged as video tutorial. Excellent introduction to KERAS and TENSORFLOW for supervised learning including image classifications. Requires an advanced computer installation and modern NIVIDA graphics cards.

Bruce, Peter and Andrew Bruce, 2017. *Practical Statistics for Data Scientists*. O'Reilly
Notes: Available online at UTD's library. Good introduction to R for data scientists. This book introduces statistical concepts relevant for machine learning. In addition, it covers several exploratory machine learning techniques.

Additional reading material will be made available as required on the course's **ELEARNING** site throughout this semester.

Software

The ***free open source*** R-environment for the operating systems

Windows, Linux and Mac OS X.
More information on the installation of Microsoft's Open R (https://mran.microsoft.com/rro/) and the development shell RStudio (https://www.rstudio.com/home/) will be provided during the first and second course week.

Assignments & Academic Calendar

[Lecture Dates, Topics, Reading Assignments, Lab and Quiz Dates]

Date	*Topic*	*Reading*	*Labs*
Aug. 19	BRAIN STORMING		*Lab01a out*
Aug. 21	INTRODUCTION AND HISTORICAL BACKGROUND		
Aug. 26	UNDERLYING MATHEMATICAL, STATISTICAL AND COMPUTATIONAL CONCEPTS (I)		
Aug. 28	UNDERLYING MATHEMATICAL, STATISTICAL AND COMPUTATIONAL CONCEPTS (II)		
Sep. 02	***LABOR DAY***		
Sep. 04	PRINCIPLE COMPONENTS I	JWHT10.2	
Sep. 09	PRINCIPLE COMPONENTS II		*Lab01b out*
Sep. 11	K-MEANS CLUSTERING	JWHT10	
Sep. 16	HIERARCHICAL CLUSTERING I		
Sep. 18	HIERARCHICAL CLUSTERING II		
Sep. 23	Classification I (Logistic Regression)	JWHT04	
Sep. 25	Classification II (Discriminant Analysis)		*Lab02 out*
Sep. 30	Resampling Methods I	JWHT05	
Oct. 02	Resampling Methods II		
Oct. 07	Moving Beyond Linearity I	JWHT07	
Oct. 09	Moving Beyond Linearity II		
Oct. 14	TREE-BASED METHODS I	JWHT08	
Oct. 16	TREE-BASED METHODS II		
Oct. 21	TREE-BASED METHODS III		*Lab03 out*
Oct. 23	SUPPORT VECTOR MACHINES I	JWHT09	
Oct. 28	SUPPORT VECTOR MACHINES II		
Oct. 30	SUPPORT VECTOR MACHINES III		
Nov. 04	NEURAL NETWORKS I	Handout	
Nov. 06	NEURAL NETWORKS II		
Nov. 11	NEURAL NETWORKS III		*Lab04 out*
Nov. 13	***GIS DAY*** & INTRODUCTION TO DEEP LEARNING	Handout	
Nov. 18	*PRESENTATIONS: SOCIAL AND ETHICAL RAMIFICATIONS OF ML*		*Project out*
Nov. 20	*PROJECT PREPARATION (I)*		
Nov. 25	***FALL BREAK***		

Nov. 27	***FALL BREAK***
Dec. 02	***FINAL EXAM***
Dec. 04	*PROJECT PREPARATION (II)*
TBA	**PROJECT PRESENTATIONS**

Labs:

Lab	*Topic*
Lab01	Prerequisites, kNN and feature transformations
Lab02	Unsupervised Learning Methods and class membership predictions
Lab03	Prediction evaluation, model refinements and tree based methods
Lab04	Support vector machines and neural networks

Course Policies

Grading (credit) Criteria

Policies:

- Labs and the final exam need to be solved ***individually*** unless otherwise stated.
- The class project can be addressed single participants and groups of maximal two members.
- **Plagiarism *cannot* be tolerated!**
- Participation is highly encourages but will not be graded.
  Note: Engagement with the course material will lead to participation and indirectly to an increased comprehension of the course material!

Requirements	*Points*
***4 Labs** @ ~12 pts: labs should be handed in as hardcopy and electronically by email.* *Note: the labs will prepare you for the final exam.*	48 pts
Presentation: Social and Ethical Ramifications of ML	4 pts
*Machine Learning **Project** with geo-referenced or socio-economic data*	24 pts
Final Exam	24 pts

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 will *not 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.

Tentative Grading Scale

Rounded Percent	Letter Grade
90-100	***A***
85-89	***A-***
80-84	***B+***
75-79	***B***
70-74	***B-***
65-69	***C+***
60-64	***C***
Below 60	***F***

These descriptions and timelines are subject to change at the discretion of the course instructor.