Course Syllabus: EE522 Selected topics in Quantitative Economics 2 (Data analytics for economics and business with Python)

1. **Lecturer and course administrator with contact information**

2. **Prerequisite:** -

3. **Class Date, Time and Venue:**

4. **Course Objectives**

The aim of the course is to provide a broad understanding of the principles and techniques of data analysis and machine learning in business and economic applications through the use of Python coding in Jupiter notebooks with intuitively visualized output. These objectives will be achieved by solving tasks/case studies and discussing theory. In this course, the real datasets from retail, marketing, social media, finance and other interesting sectors will be introduced for business decision support. For example, profiling of customers and their transactions will be used to identify the advertising strategies and create the new market segments for the business. The food reviews and customer feedbacks from social media will be analyzed by using the text mining for generating the ideas to improve its services.

On completion of the course students will be expected to:

- Understand what machine learning is and why it is essential to the design of intelligent machines.
- Understand how big data technologies can improve and extend the opportunity for its business.
- Know how to fit the suitable models for types of business decisions.
- Understand numerical computation, statistics and optimization in the context of learning.
- Have a good understanding of the business problems that arise when dealing with very small and very big data sets, and how to solve them.
- Understand the basic mathematics necessary for constructing novel data analysis and machine learning solutions.
- Be able to design and implement various data analysis and machine learning algorithms in a wide range of real-world applications.

Understanding the problems covered in the course can be important for all students seeking a career in the expanding technology-intensive field of financial engineering/Business analytics or grappling with the challenges related to the business data.

5. Course evaluation, grading criteria, grading system, and course rules

5.1. Evaluation

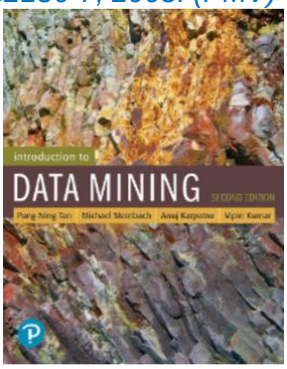
The grade breakdown is as follows:

Methods/Activities	Weights Assigned
Assignments	*30%*
Term Project	*30%*
Participation and Class Contribution/Professionalism	10%
Final Examination	*30%*
Total	100%

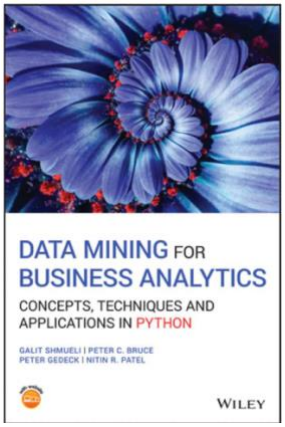
6. Reference materials

6.1. Main textbook

Pang-Ning Tan, Michael Steinbach, Vipin Kumar, ***Introduction to Data Mining*** *, Addison Wesley, ISBN: 0-321-32136-7, 2005. (PMV)*



Galit Shmueli, Peter Bruce, Inbal Yahav, Nitin Patel, and Kenneth C Lichtendahl Jr, ***Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python****, Wiley, ISBN: 978-1-119-54984-0 (GPINK)*



Foster Provost and Tom Fawcett , Data Science for Business: Fundamental principles of data mining and data analytic thinking, O'Reilly 2013. (P&F)

Ewen Gallic, Python for economists. (E)

Luis Pedro Coelho and Willi Richert, Building Machine Learning Systems with Python, 2nd Edition, Packt publishing 2013. (C&R)

6.2. Recommended texts and materials

John Guttag, Introduction to Computation and Programming using Python.

Tom M. Mitchell, Machine Learning, 1st Edition McGraw-Hill 1997.

Kevin P. Murphy. Machine Learning: A Probabilistic Perspective, MIT Press 2012.

Christopher M. Bishop. Pattern Recognition and Machine Learning, Springer 2007.

T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. Springer 2011.

S. Haykin. Neural networks and learning machines. Pearson 2008.

7. Course plan

Session	Topic	Lecturer/ Reading/Lab
	Introduction ● Why Big Data? What is data mining? Why data mining? Data Mining Process, relation to Business Intelligence techniques. ● Introduction to Data Mining Tasks (Classification, Clustering, Association Analysis, Anomaly Detection). What is a model? Basic terminologies, predictive modeling. ● Real-world data mining applications	*Ch.1(PMV)*

Session	Topic	Lecturer/ Reading/Lab
	Introduction to Python Learn the basics of Python, one of the world's most popular and powerful programming languages, and see how it can be utilized in Economics and Finance.	
	Introduction to NumPy Learn the basics of NumPy, a Python package for handling large, multi-dimensional arrays and matrices, to quicken your economic and financial analyses.	
	Data and Preprocessing ● Understanding of Data, what is data? Types of attributes, properties of attribute values, types of data, data quality ● Sampling, Data Normalization, Data Cleaning, Similarity Measures ● Feature Selection/Instance Selection, the importance of feature selection/instance selection in various big data scenarios *Introduction to Pandas*	*Ch.2(PMV)* *Basic Python: Pandas* *Case Study: analyzing stock returns* *-Bitcoin returns*

Session	Topic	Lecturer/ Reading/Lab
	Learn the basics of Pandas, a Python package for data manipulation and analysis, and discover how to access economic and financial data from APIs.	
	Visualizing Economic and Finance Data *Learn the basics of Matplotlib, a Python package for data visualization, and discover how to optimize a portfolio of stocks.*	*Case Study: analyzing asset prices* *-Stock returns*
	Statistics for Data Analysis *Learn how to calculate and interpret several descriptive statistics using the Python library NumPy.* ***Hypothesis Testing with SciPy*** *Learn SciPy, a Python module for comparative statistics, in order to perform many different statistical tests in code.*	*Case Study: COVID-19 and Resale Price of Luxury Masks* *Case Study: Firm Level Innovation and CE0 Compensation*
	Prediction and Classification: Basic Concepts and Techniques ● Basic Concepts.	*Ch.3(PMV)*

Session	Topic	Lecturer/ Reading/Lab
	● General Framework for Classification. ● Model Overfitting ●Model Selection ●Model Evaluation	
	Linear Regression	*Ch.6(GPINK)* *Case Study: 3-Factor model, 5-Factor model*
	Logistic Regression	*Ch.10(GPINK)* *Case study: Predict Loan Funding with Loan Profile*
	Classification and Regression Tree	*Ch.7(GPINK)* *Case Study: B2B customer buying stage prediction*
	Association Rules and Collaborative Filtering *Cluster Analysis*	*Ch.14,15(GPINK)*

Session	Topic	Lecturer/ Reading/Lab
		Case Study: Modeling consumer behavior for targeted marketing (banking and/or online advertising)
	Text Analytics	*Ch.20 (GPINK)* *Case Study: Sentiment Classification on movie reviews (IMDb)*
	Anomaly Detection ● Statistical-based and Density-based Methods ● Ethics of data mining, privacy, what can/do firms know?	Lecture Note
	Wrap Up *Project Presentations*	
FINAL EXAM		