Master of Science in Quantitative Finance

COURSE CODE: QF634
COURSE TITLE: APPLIED QUANTITATIVE RESEARCH METHODS

Instructor : Lim Kian Guan
Title : Professor Emeritus
Email : kglim@smu.edu.sg

PRE-REQUISITE/CO-REQUISITE/MUTUALLY EXCLUSIVE COURSE(S)
Formally none, though MQF students would have taken Quantitative Analysis of Financial Markets and Programming and Computation Financial in Term 1, so some skill sets involving Linear Regressions and some basic Python would be expected. Students from other Masters programs are also welcome provided they have similar background skill sets.

COURSE AREA
Quantitative Finance

GRADING BASIS
Graded

COURSE UNIT
1 CU

FIRST OFFERING TERM
Academic Year: AY2024-25
Academic Term: Term 1A

COURSE DESCRIPTION

This module offers students an opportunity to learn and experiment with applied quantitative research methods. There will be applications of methods to data in the context of business and finance problems. Some introductory methods including those in Machine Learning using Python packages will be taught. Technical coverages include qualitative response models, lasso and ridge regressions, regularizations, cross-validation, dimension reduction, classification methods, support vector machine, decision trees, random forest, gradient boosting, neural networks, and others. Students will practice on worked examples and then formulate, investigate, and make a final report and poster presentation on an equal-effort team applied research project based on topics covered in or directly related to the course. The instructor will provide some guidance in the formulation of the research projects – but this will be limited in order to avoid unfair advantages to groups that ask a lot more.

LEARNING OBJECTIVES

On successful completion of the course, students should be able to:

- Understand the objectives and requirements of an investigation into an applied research topic. Be able to conduct a literature review and zero in on what are potential new findings. They

should be able to collect the appropriate data set(s) for the analyses.

- Understand and be able to use appropriate research methods in machine learning to enable research findings in a rigorous way. Students will also get to understand details of how the machine works and be able to run python codes and application packages to perform prediction, analyses, and classification exercises on data sets.
- In the applied research report and poster presentations, the students should be able to explain the main ideas, data and method(s), the findings or results, and the implications. The emphasis will be on practical usages as in the workplace and less of academic publishing.

ASSESSMENT METHODS

Four take-home individual Data Analytics Exercises, each 10%	40%
Term in-class test based on materials in the Data Analytics Exercises and Lectures*	25%
Team Research Project Report and Poster Presentation in Class (2 x A3) **	30%
Class Participation	5%
Total	100%

***** Laptop computer is required for data analyses. 90 mins in-class test. First 45 mins, students use laptop and python to compute from a given data set. This part is open book. For next 45 mins, students answer questions in MCQ format on ELEARN lock-down. This second part is closed book.

** Equal marks will be allocated to each member of the same team unless there are disagreements within the team. **Teams cannot obtain assistance from instructor except some guidance in the approach to the applied research as it will pose unfair advantages. If any team has difficulty coming up with a workable project, the instructor can provide a research topic.**

Due to the short term (5 weeks) and the practice nature of the subject, there is no final sit-in examination.

ACADEMIC INTEGRITY

All acts of academic dishonesty (including, but not limited to, plagiarism, cheating, fabrication, facilitation of acts of academic dishonesty by others, unauthorized possession of exam questions, or tampering with the academic work of other students) are serious offences.

All work (whether oral or written) submitted for purposes of assessment must be the student's own work. Penalties for violation of the policy range from zero marks for the component assessment to expulsion, depending on the nature of the offence.

When in doubt, students should consult the course instructor. Details on the SMU Code of Academic Integrity maybe accessed at http://www.smuscd.org/resources.html.

ACCESSIBILITY

SMU strives to make learning experiences accessible for all. If you anticipate or experience physical or academic barriers due to disability, please let me know immediately. You are also welcomed to contact

the university's disability services team if you have questions or concerns about academic provisions: included@smu.edu.sg.

Please be aware that the accessible tables in our seminar room should remain available for students who require them.

EMERGENCY PREPAREDNESS FOR TEACHING AND LEARNING (EPTL)

Where there is an emergency that makes it infeasible to have classes on campus, classes will be conducted online via WebEx, with no disruption to the schedule. To familiarise students with the WebEx platform, part of this course maybe conducted online. The instructor will inform students of which classes, if any, will be conducted as part of this EPTL initiative.

INSTRUCTIONAL METHODS AND EXPECTATIONS

Sessions 1 through 9 will be lectures and will include about 1 hour of in-class data analytics practice with Python coding. There will also be a total of 4 graded take-home individual exercises using provided datasets. Students must submit the exercise answers individually via ELEARN for marks to be recorded.

The 10th and last session will include 1.5 hour in the first half for a term test containing python data analytics problems. The term in-class test (25%) is based on materials in the class Data Analytics Exercises and Lectures.

For the applied quantitative research project (30%), students will be divided into teams. The numbers in a team should be 4 or less. The teams must begin to think about a topic as soon as possible and start their team discussion. There will be team project report submission and also poster presentation on Session 10. All students must provide an evaluation of all projects except their own in order to obtain the 5% participation marks. See also the Housekeeping Rules at the end of this course outline.

WEEKLY LESSON PLANS

Week	TOPIC	REFERENCES
1	Session 1 Financial Portfolio Optimization	Notes in ELEARN
	Session 2 Regularization in Machine Learning Lasso and Ridge Regressions Training/Testing/Cross-Validation Hyperparameter Tuning	Notes in ELEARN Take-Home Exercise 10%
2	Session 3 Financial Reporting, Performance Metrics, Logit Regression, Dimension Reduction, PCA	Notes in ELEARN Take-Home Exercise 10%
	Session 4 Naïve Bayes, K-Nearest Neighbor Algorithms and Support Vector Machines	Notes in ELEARN
3	Session 5 Decision Trees and Random Forest	Notes in ELEARN Take-Home Exercise 10%
	Session 6 Gradient Boosting and Shap Values	Notes in ELEARN
4	Session 7 Artificial Neural Network I Multilayer Perceptron	Notes in ELEARN Take-Home Exercise 10%

	Session 8 Artificial Neural Network II RNN, LSTM, CNN	Notes in ELEARN
5	Session 9 Natural Language Processing	Notes in ELEARN
	Session 10 Test in first half followed by Team Research Project Poster Presentations in second half	Term Test 25% Applied Research 30%

RECOMMENDED TEXT AND READINGS

1. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurelien Geron, O'Reilly Publication, 2019.
2. "Practical Statistics for Data Scientists" by Peter Bruce, Andrew Bruce, and Peter Gedeck, O'Reilly Publication, 2020.

Housekeeping Rules for this course:

(1) All individual take-home exercises, term test, and project presentation reports must be submitted to ELEARN. Direct and explicit copying of fellow students' homework is not allowed. Ultimate Homework submission deadline for all Homework sets is before the start of Session 9. No marks will be awarded for nil or late submissions on ELEARN.

(2) The individual term test on the 10th session will be based on python exercises like the ones practised in class and those given in the take-home. There will also be concept tests based on the lecture materials. Mode of answering will be via MCQ format. Students must bring own computer to the classroom for the test.

(3) Team will do poster presentations in class in the second half of the 10th session. A hard printed copy of the report together with a thumb-drive containing the .csv data and ipynb file(s) must also be submitted to the instructor in class. The latter are for the instructor to check the results reported in the project. A soft copy of your research project report (pdf file) should also be uploaded to Assignments in ELEARN. (Data and ipynb files do not need to be uploaded to ELEARN.)

(4) The written applied Research report is expected to be in font size 12, single line A4 size, and typically not exceeding 15 single pages (including all tables and graphs). Names of all team members must be shown on the first page. The report should contain the main ideas and implications (5%), data and method(s) (5%), and the findings or results (20%). Original work will be given better consideration. For original work, you may use data in WRDS (available at the library using your PG account). If your team chooses to use other public data such as those in Kaggle, be careful to ensure you do not inadvertently copy ideas or results already published using those data.

(5) You can find the homeworks and the term test in the Quiz section of ELEARN at their release date/time. The password for these will be given in class. Submission is done by working on the MCQs based on the homework and test sets.

\end