H9PFDS: Programming for Financial Data Science

Module Code:	H9PFDS
Long Title	Programming for Financial Data Science APPROVED
Title	Programming for Financial Data Science
Module Level:	LEVEL 9
EQF Level:	7
EHEA Level:	Second Cycle
Credits:	5
Module Coordinator:	Rohit Verma
Module Author:	Andrea Del Campo Dugova
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	Lecturer PhD/Master's degree in a computing or cognate discipline. May have industry experience also. Tutor PhD/Master's degree in a computing or cognate discipline. May have industry experience also.

Learning Outcomes

On successful completion of this module the learner will be able to:

#	Learning Outcome Description
LO1	Investigate and evaluate key concepts and assess how to apply and utilise appropriate scientific libraries to perform computational analyses on complex datasets and practical problem domains.
LO2	Design and implement programs to solve different mathematical problems using real-world financial data.
LO3	Critically assess different data analytics approaches to apply to the data and draw appropriate conclusions from the data analytics results.
LO4	Critically assess different data analytics approaches to apply to the data and draw appropriate conclusions from the data analytics results.
LO5	Critically review and apply appropriate data mining and machine learning methods to the real-world FinTech problems

Dependencies

Module Recommendations

No recommendations listed

Co-requisite Modules

No Co-requisite modules listed

Entry requirements	Programme entry requirements must be satisfied.

H9PFDS: Programming for Financial Data Science

Module Content & Assessment

Indicative Content
Basic programming concepts Data types and structures. Loops and conditional statements. Custom functions. Structured programming.
Practical applications of programming Data input and output. Learn different file formats. Sources of data, data repositories. Understanding data characteristics (continuous, discrete, nominal, binary, structured, unstructured, time-series data) and applying manipulation techniques to data.
Libraries for scientific computing and data analysis I Advanced data manipulation techniques.
Libraries for scientific computing and data analysis II Discretisation and binning. Transformation Strategies. Scaling (normalisation, standardisation); dealing with categorical data. Feature selection.
Libraries for scientific computing and data analysis III Transformation Strategies. Dealing with outliers, Data Splitting. Dealing with missing values. Handling class imbalance.
Visualization and Exploratory Data Analysis Understand trends, outliers, and patterns in data through appropriate visualisations such as scatter plots, histograms, boxplots, pie charts, bar charts, overlayed bar charts, clustered bar charts, line charts, heatmaps, etc. Measures of central tendency (mode, median, mean) Measures of dispersion (range, variance, standard deviation)
Statistical Analysis -Hypothesis & Inference Statistical analysis, different kinds of hypothesis tests, Standard Errors Hypothesis Testing, Parametric Tests (e.g., T-Test, ANOVA, regression), Non-parametric Tests (e.g., chi-square tests) Correlation, Z-statistic, Distributions, Sample size, Confidence intervals, significance levels, p-values, effect size
Classification and Evaluation Concept of classification and its role solving real-world (FinTech) problems. Splitting a dataset, training, testing and validation, cross validation. Resampling methods. Model Evaluation. Performance measure: Accuracy, Prediction Score, Confusion Matrix, ROC curve, AUC, Precision, Recall, F1. Sample size.
Classification and Financial Text data Vectorisation. TF-IDF weighting. Sparse vectors. Dense vectors - Word Embedding Vectors.
Regression Linear Regression (LR). Multiple LR. LR for finance data. Regularisation. Evaluating regression models.
Time series Analysis Smoothing data, Analysing time series, curve fitting, seasonality. Moving averages, ARIMA (Seasonal, Non-seasonal)
Clustering K-means, Density-based clustering.

Assessment Breakdown	%
Coursework	100.00%

Assessments

Full Time

Coursework

Assessment Type:	Formative Assessment	**% of total:**	Non-Marked
Assessment Date:	n/a	**Outcome addressed:**	
Non-Marked:	Yes		

Assessment Description:
Formative assessment will be provided on the in-class individual or group activities. Feedback will be provided in written or oral format, or on-line through Moodle. In addition, in class discussions will be undertaken as part of the practical approach to learning.

Assessment Type:	Continuous Assessment	**% of total:**	20
Assessment Date:	n/a	**Outcome addressed:**	
Non-Marked:	No		

Assessment Description:
Assessment will be through an in-class, open book test, that will require learners to retrieve, extract, manipulate and present data. Learners will be also asked to make statistical inferences and draw conclusions about a population.

Assessment Type:	Project	**% of total:**	80
Assessment Date:	n/a	**Outcome addressed:**	1,2,3,4,5
Non-Marked:	No		

Assessment Description:
The terminal assessment will consist of a project that will evaluate all learning outcomes. Learners will have to identify and carry out a series of analytic tasks upon a large dataset (or a collection of datasets that are somehow related or complement each other), utilising appropriate tools and techniques for data extraction, processing, analysis and critical evaluation. The final submission will consist of an academic research paper style report as well as the implemented data analytics artefact. It is also expected students to present and communicate the results/insights of their study.

No End of Module Assessment

No Workplace Assessment

Reassessment Requirement

Coursework Only
This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.

Reassessment Description
The repeat strategy for this module is by a project that covers all learning outcomes.

H9PFDS: Programming for Financial Data Science

Module Workload

Module Target Workload Hours 0 Hours

Workload: Full Time

Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Classroom and demonstrations	24	Per Semester	2.00
Tutorial	Mentoring and small-group tutoring	12	Per Semester	1.00
Tutorial	Independent learning	89	Per Semester	7.42
			Total Weekly Contact Hours	10.42

Workload: Blended

Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Classroom and demonstrations	12	Per Semester	1.00
Tutorial	Mentoring and small-group tutoring	12	Per Semester	1.00
Directed Learning	Directed e-learning	12	Per Semester	1.00
Independent Learning	Independent learning	89	Per Semester	7.42
			Total Weekly Contact Hours	3.00

Workload: Part Time

Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	Classroom and demonstrations	24	Per Semester	2.00
Independent Learning	Independent learning	89	Per Semester	7.42
Tutorial	Mentoring and small-group tutoring	12	Per Semester	1.00
			Total Weekly Contact Hours	3.00

Module Resources

Recommended Book Resources

John D. Kelleher,Brian Mac Namee,Aoife D'Arcy. (2020), Fundamentals of Machine Learning for Predictive Data Analytics, second edition, MIT Press, p.853, [ISBN: 978-0262044691].

Hastie, T., Tibshirani, R., & Friedman, J.. (2016), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Ed. Springer.

Aurelien Geron. (2019), Hands-On Machine Learning with Scikit-Learn, Keras & Technology, 2nd ED. O'REILLY.

Supplementary Book Resources

Kevin P. Murphy. (2012), Machine Learning: A Probabilistic Perspective, MIT Press, p.1102, [ISBN: 978-0262018029].

Ian Goodfellow,Yoshua Bengio,Aaron Courville. (2016), Deep Learning, MIT Press, p.801, [ISBN: 978-0262035613].

James, G., Witten, D., Hastie, T., & Tibshirani, R.. (2017), An Introduction to Statistical Learning, Springer, p.426, [ISBN: 978-1461471370].

Max Kuhn,Kjell Johnson. (2013), Applied Predictive Modeling, Springer, p.600, [ISBN: 978-1461468486].

Shai Shalev-Shwartz,Shai Ben-David. (2014), Understanding Machine Learning, Cambridge University Press, p.415, [ISBN: 978-1107057135].

John Hearty. (2016), Advanced Mastering Learning with Python, Packt Publishing, p.278, [ISBN: 978-1784398637].

This module does not have any article/paper resources

Other Resources

[Website], Machine Learning Stanford,
https://www.coursera.org/course/ml

[Website], DataCamp,
http://www.datacamp.com

[Website], UCI Repository,
http://www.ics.uci.edu/~mlearn/MLReposit ory.html

[Website], WEKA,
http://www.cs.waikato.ac.nz/ml/weka/

[Website], Kaggle Competitions,
https://www.kaggle.com/competitions

Discussion Note:	