SEMESTER 1 2022/23

UNIVERSITY OF Southampton

COURSEWORK BRIEF:

Module Code:	MANG6556	Assessment:	Individual Coursework	Weighting:	100%
Module Title:	Credit Risk & Data Analytics				
Module Leader:	Dr Huan Yu				
Submission Due Date: @ 16:00	13 Jan 2023, 16:00		Word Count:		2500
Method of Submission:	Electronic via Blackboard Turnitin ONLY (Please ensure that your name does not appear on any part of your work)				

Any submitted after 16:00 on the deadline date will be subject to the standard University late penalties (see below), unless an extension has been granted, in writing by the Senior Tutor, in advance of the deadline.

University Working Days Late:	Mark:
1	(final agreed mark) * 0.9
2	(final agreed mark) * 0.8
3	(final agreed mark) * 0.7
4	(final agreed mark) * 0.6
5	(final agreed mark) * 0.5
More than 5	0

This assessment relates to the following module learning outcomes:

A. Knowledge and Understanding	A1. Understand the potential of CRISP-DM and data analytics, particularly in the retail lending sector. A2. Demonstrate a critical understanding of different types of data analytics methods and the problems they can solve. A3. Interpret the output of statistical techniques used for the main data analytics applications.
B. Subject Specific Intellectual and Research Skills	B1. Identify the statistical models appropriate for analysing the various decisions that confront a data analyst in different industries. B2. Work with software to develop data analytics solutions, such as predictive scorecards, clustering models, and different types of regressions. B3. Assess the relevance of statistical package outputs to the decisions being addressed.
C. Transferable and Generic Skills	C1. Critically analyse practical difficulties that arise when implementing retail credit risk models; understand the cross-fertilisation potential to other business contexts (e.g., fraud detection, marketing, CRM, etc.). C2. Demonstrate an ability to use world-class software and to interpret its output in the relevant techniques. C3. Manage time and tasks effectively in the context of individual study.

Coursework Brief:

Question 1 (60 marks)

The dataset 'Credit data.xlsx' contains data on 10,000 borrowers and whether they subsequently

experienced serious delinquency (see variable 'SeriousDlqin2yrs'). Assume the lender now wishes to use this data to build a credit scoring model that predicts serious delinquency based on the other variables. The dataset contains the following variables:

Variable Name	Description
SeriousDlqin2yrs	**Person experienced 90 days past due delinquency or worse**
RevolvingUtilizationOfUnsecuredLines	Total balance on credit cards and personal lines of credit except real estate and no installment debt like car loans divided by the sum of credit limits
age	Age of borrower in years
NumberOfTime30-59DaysPastDueNotWorse	Number of times borrower has been 30-59 days past due but no worse in the last 2 years.
DebtRatio	Monthly debt payments, alimony,living costs divided by monthy gross income
MonthlyIncome	Monthly income
NumberOfOpenCreditLinesAndLoans	Number of Open loans (installment like car loan or mortgage) and Lines of credit (e.g. credit cards)
NumberOfTimes90DaysLate	Number of times borrower has been 90 days or more past due.
NumberRealEstateLoansOrLines	Number of mortgage and real estate loans including home equity lines of credit
NumberOfTime60-89DaysPastDueNotWorse	Number of times borrower has been 60-89 days past due but no worse in the last 2 years.
NumberOfDependents	Number of dependents in family excluding themselves (spouse, children etc.)

1.1 Carefully pre-process the dataset by considering the following activities: (30 marks)

- Exploratory data analysis.
- Missing value handling (if any), including a suitable analysis of missing values and justification of the chosen method.
- Outlier detection and treatment (if any), with appropriate analysis/justification.
- Binning the variables (if deemed useful)
- Coding the discrete variables using Weights of Evidence.
- Splitting the data set into a training and test set.

1.2 Build an intuitive and predictive scorecard using a logistic regression classifier and report the following: (30 marks)

- The most important variables
- The impact of the variables on the target
- The performance of the model. Use various performance metrics and discuss their relationship if any.
- Compare this scorecard with the result of a Random Forest model run over the data. Discuss your results.
- Why do banks typically use Logistic Regression as their base classifier? What do banks win and lose by doing this?

Carefully report the various steps of your methodology and discuss your results in a rigorous way!

NOTE: It is unlikely that different students will come up with the exact same parameter estimates. Special consideration will be given to submissions whose estimates are identical.

Question 2 (20 marks)

Find an academic paper published in 2019 or later (based on online or print publication date) discussing a real-life
application of credit risk or data analytics. It is important that the dataset analysed in the paper consists of real-life
(Not artificial) data. The publication outlets in which to look for a paper are:

- Management Science
- Operations Research
- INFORMS Journal on Computing
- INFORMS Journal on Applied Analytics
- Journal of Machine Learning Research
- European Journal of Operational Research
- Production and Operations Management

- Manufacturing & Service Operations Management
- ICDM (The IEEE International Conference on Data Mining)
- NeurIPS (Conference on Neural Information Processing Systems)
- KDD (ACM SIGKDD Conference on Knowledge Discovery and Data Mining)

The other journals which are not on the list are not acceptable.

Once you have found an appropriate paper, report the following in separate subsections:

- Title, authors, and complete citation (e.g., journal name, volume/issue, year, …)
- The data mining problem considered in the paper
- The data mining methodology used in the paper
- The results reported in the paper
- A critical discussion of the model and results (assumptions made, shortcomings, limitations, …).

Make sure you demonstrate that you understand what the article is all about and are able to provide a critical
discussion.

Do not copy and paste from the article. Using Turnitin, this will be easily detected!

Question 3 (20 marks)

Apply the methodology you reviewed in Question 2 into the dataset 'Credit data.xlsx' provided in Question 1 and report the following:

- The analytic steps of applying the reviewed methodology in credit data
- The performance of the reviewed methodology for credit data
- The business implications of using the reviewed methodology on credit data

NOTE: The methodology should be different from methods applied in Question 1.

SEMESTER 1 2022/23

UNIVERSITY OF Southampton

Nature of Assessment: This is a SUMMATIVE ASSESSMENT. See 'Weighting' section above for the percentage that this assignment counts towards your final module mark.

Word Limit: +/-10% either side of the word count (see above) is deemed to be acceptable. Any text that exceeds an additional 10% will not attract any marks. The relevant word count includes items such as cover page, executive summary, title page, table of contents, tables, figures, in-text citations and section headings, if used. The relevant word count excludes your list of references and any appendices at the end of your coursework submission.
You should always include the word count (from Microsoft Word, not Turnitin), at the end of your coursework submission, before your list of references.

Title/Cover Page: You must include a title/ cover page that includes: your Student ID, Module Code, Assignment Title, Word Count. This assignment will be marked anonymously, please ensure that your name does not appear on any part of your assignment.

References: You should use the Harvard style to reference your assignment. The library provide guidance on how to reference in the Harvard style and this is available from: http://library.soton.ac.uk/sash/referencing

Submission Deadline: Please note that the submission deadline for Southampton Business School is 16.00 for ALL assessments.

Turnitin Submission: The assignment MUST be submitted electronically via Turnitin, which is accessed via the individual module on Blackboard. Further guidance on submitting assignments is available on the Blackboard support pages.

It is important that you allow enough time prior to the submission deadline to ensure your submission is processed on time as all late submissions are subject to a late penalty. We would recommend you allow 30 minutes to upload your work and check the submission has been processed and is correct. Please make sure you submit to the correct assignment link.

Email submission receipts are not currently supported with Turnitin Feedback Studio LTI integrations, however following a submission, students are presented with a banner within their assignment dashboard that provides a link to download a submission receipt. You can also access your assignment dashboard at any time to download a copy of the submission receipt using the receipt icon. It is vital that you make a note of your Submission ID (Digital Receipt Number). This is a unique receipt number for your submission, and is proof of successful submission. You may be required to provide this number at a later date. We recommend that you take a screenshot of this page, or note the number down on a piece of paper.

The last submission prior to the deadline will be treated as the final submission and will be the copy that is assessed by the marker.

It is your responsibility to ensure that the version received by the deadline is the final version, resubmissions after the deadline will not be accepted in any circumstances.

Important: If you have any problems during the submission process you should contact ServiceLine immediately by email at Serviceline@soton.ac.uk or by phone on +44 (0)23 8059 5656.

Late Penalties: Further information on penalties for work submitted after the deadline can be found here.

Special Considerations: If you believe that illness or other circumstances have adversely affected your academic performance, information regarding the regulations governing Special Considerations can be accessed via the Calendar: http://www.calendar.soton.ac.uk/sectionIV/special-considerations.html

Extension Requests: : Extension requests along with supporting evidence should be submitted to the Student Office as soon as possible before the submission date. Information regarding the regulations

governing extension requests can be accessed via the Calendar: http://www.calendar.soton.ac.uk/sectionIV/special-considerations.html

Academic Integrity Policy: Please note that you can access Academic Integrity Guidance for Students via the Quality Handbook: http://www.southampton.ac.uk/quality/assessment/academic_integrity.page?. Please note any suspected cases of Academic Integrity will be notified to the Academic Integrity Officer for investigation.

Feedback: Southampton Business School is committed to providing feedback within 4 weeks (University working days). Once the marks are released and you have received your feedback, you can meet with your Module Leader / Module Lecturer / Personal Academic Tutor to discuss the feedback within 4 weeks from the release of marks date. Any additional arrangements for feedback are listed in the Module Profile.

Student Support: Study skills and language support for Southampton Business School students is available at: http://www.sbsaob.soton.ac.uk/study-skills-and-language-support/.

External Examiner:	
External Examiner Comments:	
Final Approval by External Examiner Date:	
Module Leader Response to External Examiner: (Please note these comments are REQUIRED and will be sent to the External Examiner)	