

Solent University Coursework Assessment Brief

Assessment Details

Module Title:	Machine Learning
Module Code:	COM624
Module Leader:	Dr. Bacha Rehman
Level:	Year 3 (FHEQ Level 6)
Assessment Title:	Software Development Project
Assessment Number:	AE1
Assessment Type:	Software Artefact and Report
Restrictions on Time/Word Count:	3000 words
Consequence of not meeting time/word count limit:	There is no penalty for submitting above or below the word/count limit, but students should be aware that there is a risk they may not maximise their potential mark.
Individual/Group:	Individual
Assessment Weighting:	100%
Issue Date:	Week commencing 02/10/2023
Hand In Date:	12/01/2024 at 4 PM
Planned Feedback Date:	February 2024
Mode of Submission:	On-line
Anonymous Marking	This assessment: Is exempt from anonymous marking.

Assessment Task

The tutor provided case scenario is as follow however, students can provide their own case scenario so far, it meets the required standard similar to the provided data and scenario (i.e., students' provided scenario and dataset should be agreed and sign-off by the unit tutor from the onset).

Case Scenario

Solent Financial Technology (SOLFINTECH) is a leading financial multinational organisation that deals with stocks and shares, savings and investments. The organisation operates an online investment platform which accommodates over 50 million subscribers (clients) with over 150 billion pounds worth of investments. SOLFINTECH is a secure organisation that trades on multiple stock exchange platforms such as FTSE 100 and equities. Due to the large customer base and competition between the newly spring up financial tech organisations, the company intends to implement an Intelligent Stock Trader (IST) platform. The IST platform will perform intelligent trading by purchasing FTSE company stocks or equities at lower prices and sell at higher prices. The platform will perhaps anticipate stock/equities prices on a daily, weekly, monthly and the maximum of quarterly basis. (i.e., the system will identify stocks with the potential of buying low and selling high at the specified interval). That said, a registered user may ask to know the stock that will provide a specified amount of profit at a specified interval. For example, the user may want to know a transaction(s) that will fetch £50 profit today or £500 in profit a month. Therefore, the system should suggest a particular stock or group of stocks that will meet the set profit margin.

The correctness of the system's prediction will be tested with real-world data and you are expected to demonstrate substantial percentage correctness. For example, the system's prediction for yesterday can be validated with today's data, and the prediction for last week could be validated with this week's data etc.

Your task

You have been employed as a solution architect, with expertise in Artificial Intelligence (machine learning, precisely) to develop a dummy system that will allow the company to test and evaluate the sustainability of this intended platform. You are free to utilise data of any existing stock exchange platform such as YahooFinance or GOOGLFINANCE among other systems whether real-time or persistent. You may even consider the user and the type of stock of interest.

The activity entails the development of a software model, i.e., a visualised decision-making system for stock and shares investment with a written individual report that underpins and evaluates the development process.

The visualised system is expected to suitably help test for the company's intended recommendations but not limited to the following requirements.

System requirements

- The system should use appropriate data visualisation approach such as, graphs and chart to aid intuitive data representation.

- The system should be a decision-making support system that will assist users make decisions based on the intended scenario, for instance, users may specify the interval (day or week) or profit or stock of interest.
- The design and development of the system must evidence data connections (you may use a real-world data platform such as, YahooFinance or GOOGLFINANCE).
- The system may be a standalone or web-based application, which displays graphical performance and trends for each of the key areas such as time against amount.
- You will be expected to apply suitable Machine Learning techniques and algorithms with justification of their suitability for the system. You may combine multiple techniques and algorithms to evidence sophistication.
- The system should be interactive and able to handle real-time information, i.e., any alteration to the data from the interface must reflect immediately.
- The model is expected to address all areas of business operations determinable from the dataset and case scenario.
- The model may be able to integrate data systems for historic database transactional data, forecast models and real-time data systems as available from the internet.
- The system may also allow the users plan and forecast investment using a range of scenarios, i.e., the system permits alteration to answer 'What-If' questions.

Data

To develop efficient model, a substantive amount of data which is exemplary of the intended prediction is required. Thus, you are required to utilise data from any stock exchange platform to develop your system. It is recommended that you utilise stock and shares data of around 100 companies with starting/opening, closing, high and low-price data among other variables to effectively demonstrate sophistication. You may want to utilise real-time data connection or persistence data source.

Software

The choice of suitable machine learning models or algorithms for the development of the software application is at the student's discretion, but as a minimum, the system MUST consist of a data source (persistent or real-time), which the graphically outputted data is sourced. Also, the choice of a front-end development is at the discretion of the student, but an elegant dashboard offering a reliable data visualisation is expected. Student should utilise/build on the IT skills already developed on the course to develop a concrete front and back system. Students are free to utilise any comfortable programming language (HTML, Python, Java and JavaScript among others), particularly for the front-end of the application.

Finally, it is expected that students utilise a specific development methodology for designing, implementing, testing and managing the software (this should include a set of user instructions if these have not been included in any 'Help' menus).

Interface

A system interface which improves user's interaction with the developed system. The system interface must gear towards the system's needs/requirements, allowing users to achieve optimal interaction with the system. Therefore, the system **MUST** be intuitive, i.e., simple to operate.

Report

You are required to produce a formal technical report of approximately 3000 words that critically evaluates the developed model/system (i.e., explaining the process of development). In critically evaluating the developed system, emphasis should be laid on the application of existing theoretical models and suitability where appropriate. Evidence of understanding through critiques of existing concepts in literature **MUST** comprehensively reflect in the report. Also, any adopted model should be well justified, particularly with regard to their suitability for the problem case, the challenges you undertake and the issues concerning its appropriateness.

You will be expected to discuss data dimensions, pre-processing, model(s) suitability and system development; your understanding of machine learning principles and concepts and its applications for system development and decision-making will be examined.

Appendices

While the contents of appendices do not count towards the word limit, the section may include detailed evidence of the models, functionality and the process of development. This could include images, system architecture and development guides among other things. It may also contain bibliographies and references of texts. All materials should be properly cited in text and referenced as appropriate.

Demonstration and Presentation

To demonstrate an understanding of machine learning principles and concepts, and the process of development, individual exhibition/presentation slot shall be communicated by the tutor as appropriate. Students may prepare a PowerPoint presentation to discuss problem case analysis, challenges, model evaluation and process of development with an exhibition of the developed system. The demonstration is equally a critical aspect of the assessment; thus, it is important to attend this session as it aims to help you and your tutor have a better understanding of the submitted portfolio. Your tutor shall provide feedback in this session to allow you to reflect on your work and in developing better systems and conducting effective research in the future.

System details with tasks specified:

- 1. Get Data:** Use any financial data provider for NASDAQ-100-Inex, such as <https://finance.yahoo.com/quote/%5ENDX/> . You can see the list of 100 NASDAQ companies at <https://www.slickcharts.com/nasdaq100> . Search each company on online stock data provider, such as Yahoo Finance etc. and download at least 1-year data. You should ideally use python script/library to download the data and get the live data stream for interactive charts.

2. **Grouping:** Apply Clustering to group the 100 indexes companies into 4 groups.

Hints:

- a. 100 stocks as rows
- b. Around 52 (weeks) x 5 (days per week) = 260 columns
- c. Use dimensionality reduction algorithm, such as PCA and LDA etc. to reduce 260 columns to around 10 or any other number of columns you think are more suitable for each stock.
- d. Use any clustering algorithm (k-mean, hierarchical or DB-scan etc.) to group stocks together. I will suggest using k-mean with number of clusters = 4. However, if you think that another clustering algorithm provide better grouping/clustering then do mention it in your report for extra credits.
- e. You got 4 groups of stocks.
- f. Now select any one stock of your choice per group.
- g. These are your company/stock to analyse. These stocks represent their respective group/category.
- h. **Warning: Students are expected to select different stocks from their respective stocks' groups. Two or more students found to have exactly same 4 stocks selected for their project will need to explain about the reason.**

3. **Correlation:** Present Top-10 Highly Correlated (Positive and Negative) for the 4 stocks selected.

4. **EDA:** Perform Exploratory Data Analysis for each selected stock, i.e. explore temporal structure, visualize the distribution of observation, and investigate the change in distribution over intervals. You can present extra EDA analysis for extra credit/grades.

5. **ML Models for Prediction and Forecasting:** Propose at least 4 machine learning models, for example ARIMA, LSTM, Any Regressor, and/or Facebook Prophet model etc., and train it over each of the stock data to **predict and forecast** the price action. Critically analyse and compare all 4 models. You can train more than 4 models for extra credits/grades.

6. **General Trading Signals and Analysis:** Generate Buy and Sell Signal based on the forecast you are having for next week, 2-weeks, and month. You can provide extra features such as trend, embedding financial news, indicators etc. for extra credits.

7. **GUI:** User should be able to perform these tasks (EDA, Correlation, Prediction, Forecasting and generating buy/sell signals etc.) and analysis using a graphical user interface provided. More appealing and detailed GUI will get extra credits.

Assessment Criteria

CRITERIA	Upper FIRST A1 -A2 Exceeds Expectation in many aspects		FIRST A3 -A4 Substantially exceeds expectations		UPPER SECOND B1 - B3 (High) Meets learning outcomes and exceeds expectations in several aspects			LOWER SECOND C1- C3 (Good) Meets learning outcomes and sometime exceeds expectations			THIRD D1 - D3 (Competent) Meets learning outcomes			FAIL F1 - F3 (incomplete/poor) Fails to meet learning outcomes		
Solent Grade	A1	A2	A3	A4	B1	B2	B3	C1	C2	C3	D1	D2	D3	F1	F2	F3
DATA COLLECTION AND USAGE	Excellent data integration from sources both internal and external to organisation.		Excellent data integration from sources.		Data used represents real world reality and information created appropriate			Data used appropriate.			Simple appraisal of academic and commercial material – meets threshold requirements			Poor and insufficient understanding of academic and commercial material and requirements		
System design and development	Excellent academic theory researched and applied in many ramifications design of new model. Implementation matches design.		Academic theory researched and applied in design of new model. Implementation matches design.		Models suitable for commercial use. Appropriate use of graphic display. Implementation matches design, or if not, the reasons for this are explained			Models represent real world reality. Little evidence of use of the more advanced implementation technologies covered in the module			Models used quite basic and do not fully cover key business areas Implementation matches design, or if not, the reasons for this are explained clearly in the write-up			Poor understanding of requirements. Weak Models. A minimal effort; up to one use case successfully implemented		
Ability to handle complex questions with robust interface	Excellent application of academic theory applied in design. Good articulation and identification of KPI's Robust error handling in many aspects and a user-friendly interface.		Academic theory applied in design. Good articulation and identification of KPI's Robust error handling and a user-friendly interface.		Key business issues resolved. There may be room for improvement in your error handling. Some evidence of use of the more advanced implementation technologies covered			Relationship between cause and effect demonstrated. Little error handling.			In Simple What-if scenarios implemented. Minimal or no error handling process.			Incomplete design and implementation. Trivial 'what-if' scenarios. No interface has been presented.		

			in the module. referencing and formatting.			
Evaluation of the system	Critical evaluation of process and system. Further areas for improvement explored beyond expectation. Excellent demonstration of taught concepts and beyond. A robust presentation.	Critical evaluation of process and system. Further areas for improvement explored. Excellent demonstration of taught concepts. A robust presentation.	Full evaluation of suggested system. New ideas offered to acquire, collect, store and use data. Very good demonstration. The presentation slides cover expected concepts	Suggested solution meets needs of all stakeholders. Good demonstration of software artefact and presentation is clear. Core concepts are covered in the presentation	Threshold decision making methods used. Acceptable demonstration of understanding of taught concepts with necessary principles covered in presentation	Simple models suggested, do not meet company needs. Presentation slides do not present appropriate context. Poor or no presentation or demonstration

Learning Outcomes

This assessment will enable you to demonstrate in full or in part your fulfilment of the following learning outcomes identified in the Module Descriptor:

- Demonstrate significant understanding of the basics and underlying challenges of machine learning problems in the context of data models, selecting appropriate models and level of complexity among others.
- Draw a comprehensive knowledge of the strengths and weaknesses of popular machine learning techniques and their applications through research about data and machine learning protocols.
- Relate with the underlying mathematical concepts and corresponding relationships within machine learning algorithms particularly, linear non-linear models.
- Evaluate and justify the suitability of machine learning types including supervised and unsupervised learning to implement models for language, vision, speech and novel decision making.
- Design and develop various machine learning models/systems to solve a range of real-world problems.
- Develop an understanding of machine learning conflicts with legal, ethical and societal issues.

Living CV

As part of the University's Work Ready, Future Ready strategy, you will be expected to build a professional, Living CV as you successfully engage and pass each module of your degree.

The Living CV outputs evidenced on completion of this assessment are as presented in the learning outcomes.

Please add these to your CV via the Living CV builder platform on Solent Futures Online [Solent Futures Online](#)

Important Information

Late Submissions

You are reminded that:

- i. If this assessment is submitted late i.e., within 7 calendar days of the submission deadline, the mark will be capped at 40% if a pass mark is achieved;
- ii. If this assessment is submitted later than 7 calendar days after the submission deadline, the work will be regarded as a non-submission and will be awarded a zero;
- iii. If this assessment is being submitted as a referred piece of work, then it must be submitted by the deadline date; any Refer assessment submitted late will be regarded as a non-submission and will be awarded a zero.

Assessment regulations

Extenuating Circumstances

The University's Extenuating Circumstances (EC) procedure is in place if there are genuine short term exceptional circumstances that may prevent you submitting an assessment. If you are not 'fit to study', you can either request an extension to the submission deadline of 7 calendar days or you can request to submit the assessment at the next opportunity, i.e., the resit period (as a Defer without capping of the grade). In both instances you must submit an EC application with relevant evidence. If accepted under the university regulations, there will be no academic penalty for late submission or non-submission dependent on what is requested. You are reminded that EC covers only short-term issues (20 working days) and that if you experience longer term matters that impact on your learning then you must contact the Student Hub for advice.

Please find a link to the EC policy below:

Extenuating Circumstances

Academic Misconduct

Any submission must be your own work and, where facts or ideas have been used from other sources, these sources must be appropriately referenced. The University's Academic Handbook includes the definitions of all practices that will be deemed to constitute academic misconduct. You should check this link before submitting your work.

Procedures relating to student academic misconduct are given below:

Academic Misconduct

Ethics Policy

The work being carried out must be in compliance with the university Ethics Policy. Where there is an ethical issue, as specified within the Ethics Policy, then you will need an ethics release or ethics approval prior to the start of the project.

The Ethics Policy is contained within Section 2S of the Academic Handbook:

Ethics Policy

Grade marking

The University uses an alpha numeric grade scale for the marking of assessments. Unless you have been specifically informed otherwise your marked assignment will be awarded a letter/number grade. More detailed information on grade marking and the grade scale can be found on the portal and in the Student Handbook.

Grade Marking Scale

Guidance for online submission through Solent Online Learning (SOL)

Online Submission