

Unit 26: Machine Learning

Authorised Assignment Brief 1

Student Name/ ID Number	
Unit Number and Title	26: Machine Learning
Academic Year	
Unit Tutor	
Assignment Title	Disease Detection of Fruits Using Machine Learning Techniques
Issue Date	
Submission Date	
Submission Format	
Submission 1 A report of 2000 to 2500 words, detailing various different machine learning techniques that can be used given the scenario, to improve the products efficacy.	
Submission 2 A complete machine learning solution to a problem of your choice from any of the kaggle datasets. The solution should contain a source code using the chosen dataset, a manual describing the various aspects and modules of the source code, and a trained model. The source code should structured and should contain comments, according to industry standards. The manual should contain well organized headings and screenshots, to appropriately describe your code and show its working. The manual should be between 1000 to 2000 words, with an emphasis on diagrams and screenshots.	

Unit Learning Outcomes
<p>L01 Analyse the theoretical foundation of machine learning to determine how an intelligent machine works</p> <p>L02 Investigate the most popular and efficient machine learning algorithms used in industry</p> <p>L03 Develop a machine learning application using an appropriate programming language or machine learning tool for solving a real-world problem</p> <p>L04 Evaluate the outcome or the result of the application to determine the effectiveness of the learning algorithm used in the application</p>
Transferable skills and competencies developed
<p>Computing-related cognitive skills</p> <ul style="list-style-type: none"> • Demonstrate knowledge and understanding of essential facts, concepts, principles and theories relating to computing and computer applications • Use such knowledge and understanding in the modelling and design of computer-based systems for the purposes of comprehension, communication, prediction and the understanding of trade-offs • Recognise and analyse criteria and specifications appropriate to specific problems, and plan strategies for their solutions

- Analyse the extent to which a computer-based system meets the criteria defined for its current use and future development
- Deploy appropriate theory, practices and tools for the design, implementation and evaluation of computer-based systems.

Computing-related practical skills

- The ability to evaluate systems in terms of quality attributes and possible trade-offs, presented within the given problem
- The ability to plan and manage projects to deliver computing systems within constraints of requirements, timescale and budget
- The ability to recognise any risks and safety aspects that may be involved in the deployment of computing systems within a given context
- The ability to deploy effectively the tools used for the construction and documentation of computer applications, with particular emphasis on understanding the whole process involved in the effective deployment of computers to solve practical problems
- The ability to critically evaluate and analyse complex problems, including those with incomplete information, and devise appropriate solutions, within the constraints of a budget.

Generic skills for employability

- Intellectual skills: critical thinking; making a case; numeracy and literacy
- Self-management: self-awareness and reflection; goal setting and action planning, independence and adaptability; acting on initiative; innovation and creativity
- Contextual awareness, e.g. the ability to understand and meet the needs of individuals, business and the community, and to understand how workplaces and organisations are governed.

Vocational scenario

You are a Machine Learning Researcher at AgriTech Innovations, a company dedicated to advancing agricultural technology. Your latest project involves working with FreshFarms, a large-scale fruit farming cooperative, to develop a machine learning system for early detection and diagnosis of fruit diseases.

Background:

FreshFarms grows a variety of fruits, including apples, oranges, grapes, and strawberries. Disease outbreaks in their orchards and vineyards can lead to significant crop losses and affect fruit quality. By implementing a machine learning-based solution, FreshFarms aims to detect diseases early, reduce the use of pesticides, and improve overall yield and fruit quality.

Literature Review:

1. Conduct a comprehensive literature review on the application of machine learning in agriculture, focusing on fruit disease detection and diagnosis.
2. Summarize key findings and identify state-of-the-art techniques and algorithms relevant to the project.

Data Collection and Preprocessing:

1. Collaborate with FreshFarms to collect a large dataset of images showing healthy and diseased fruits and foliage.
2. Preprocess the images to enhance quality, annotate the data with disease labels, and create a balanced dataset.
3. Perform exploratory data analysis (EDA) to understand the distribution of diseases and identify any patterns.

Model Development:

1. Develop machine learning models using techniques for image recognition and classification.
2. Compare the performance of different models using appropriate metrics such as accuracy, precision, recall, and F1-score.

Model Validation and Testing:

1. Split the data into training, validation, and test sets.
2. Validate the models using cross-validation techniques to prevent overfitting.
3. Test the final models on the test set to evaluate their performance and ensure they generalize well to new images.

Implementation and Deployment:

1. Develop a mobile application or web interface that allows farmers to upload images of fruits and foliage.
2. Integrate the machine learning models into the application to provide real-time disease detection and diagnosis.
3. Ensure the implementation is scalable and can handle high volumes of image uploads.

Ethical Considerations and Bias Mitigation:

1. Analyze the ethical implications of using machine learning in agriculture.

2. Identify potential biases in the dataset and models, and implement strategies to mitigate these biases.
3. Ensure the models are fair and do not disadvantage any specific types of fruit or growing conditions.

Reporting and Documentation:

1. Document the entire process, including data collection, preprocessing steps, model development, validation, and deployment.
2. Prepare a comprehensive report summarizing the findings, model performance, and potential impact on disease management and crop yield.

Present the report to FreshFarms stakeholders, highlighting key insights and recommendations for future work.

Assignment activity and guidance

Activity 1

Your report needs detail various machine learning techniques that can be potentially used to improve your clients services. You will need to explore the various generic requirements of such a company, and different ways that machine learning techniques can be used.

Essentially, you need to understand how machine learning can be used in general, along with how it can be used as a supplementary technology to assist in an IT products function. Research and consider standard operating usecases. After detailing the expected usecases of such a company, propose different machine learning algorithms that can be used to manage the company's operations by directly discussing their part in the expected usecases. You may also detail different techniques to optimize the aforementioned algorithms.

Investigate the different algorithms, compare them, and though working examples, demonstrate their efficacy. Prove the effectiveness of using machine learning to develop smart solutions.

Activity 2

You need to develop a machine learning solution for a kaggle dataset. First, visit <https://www.kaggle.com/datasets>, and download a dataset of your choice. The site should provide various important details regarding your chosen dataset, such as the problem being solved, the recommended suggestions, and what to expect from the dataset.

After downloading the dataset, you need to decide what algorithm to use to solve the problem defined. After coming to a decision, begin coding your solution using python as a programming language. You may use built-in libraries, or third-party open-source libraries.

You need to code your way to first access the data, read and pre-process it, and finally ready it for computation.

The machine learning algorithm you have chosen needs to be implemented, and fed the readied data. Now you need to train your model, and generate results.

Test your results to see if they are accurate and precise, and within the required margin of error. If not, then optimize your model, and retrain. Continue till your computed results are inline with expected results.

Finally, create a manual to explain the various aspects of the code and the results produced, including insights into your decision making process

during the development phase. Include screenshots to showcase the running example of your model. In the manual, you need to show to what degree your implemented solution can solve the problem defined, and suggest potential improvements for future development. Discuss issues that you faced. Evaluate the effectiveness of your implementation and the results it has produced, and use said evaluation to conclude the efficacy of the implemented techniques.

Learning Outcomes and Assessment Criteria

Pass	Merit	Distinction
LO1 Analyse the theoretical foundation of machine learning to determine how an intelligent machine works		D1 Critically evaluate why machine learning is essential to the design of intelligent machines.
P1 Analyse the types of learning problems. P2 Demonstrate the taxonomy of machine learning algorithms.	M1 Evaluate the category of machine learning algorithms with appropriate examples.	
LO2 Investigate the most popular and efficient machine learning algorithms used in industry		
P3 Investigate a range of machine learning algorithms and how these algorithms solve learning problems. P4 Demonstrate the efficiency of these algorithms by implementing them using an appropriate programming language or machine learning tool.	M2 Analyse these algorithms using an appropriate example to determine their power.	

Pass	Merit	Distinction
LO3 Develop a machine learning application using an appropriate programming language or machine learning tool for solving a real-world problem		D2 Critically evaluate the implemented learning solution and its effectiveness in meeting end user requirements.
P5 Prepare training and test data sets in order to implement a machine learning solution for an appropriate learning problem. P6 Implement a machine learning solution with a suitable machine learning algorithm and demonstrate the outcome.	M3 Test the machine learning application using a range of test data and explain each stage of this activity.	
LO4 Evaluate the outcome or the result of the application to determine the effectiveness of the learning algorithm used in the application		
P7 Discuss whether the result is balanced, underfitting or overfitting. P8 Analyse the result of the application to determine the effectiveness of the algorithm.	M4 Evaluate the effectiveness of the learning algorithm used in the application.	

