Study Guide

Machine Learning 382

Academic Year 2024

"Research has shown that it takes 31 days of conscious effort to make or break a habit. That means, if one practices something consistently for 31 days, on the 32nd day it does become a habit. Information has been internalized into behavioral change, which is called transformation."

Shiv Khera

Academic Year 2024

Version 2.0 (January 2024)

TABLE OF CONTENTS

MODULE DESCRIPTION	
Module Name	Machine Learning 382
Module Code	MLG382
Qualification	B.Comp
NQF Level	8
Duration (weeks)	3
Pre-requisites	None

OUTCOMES

Purpose

Machine learning is a subset of artificial intelligence. The overall purpose of the program is to produce graduates who can think clearly and critically and apply the knowledge of statistical techniques to give computers the ability to progressively improve performance on a specific task with data, without being explicitly programmed.

Outcomes

Upon successful completion of this module, the student will be able to:

- Demonstrate an integrated knowledge of the central areas of machine learning, including an understanding of and the ability to apply and evaluate the key terms, concepts, facts, principles, rules and theories of machine learning; and detailed knowledge of an area or areas of specialisation and how that knowledge relates to other fields, disciplines or practices.
- Demonstrate an understanding of knowledge as contested and the ability to evaluate types of knowledge and explanations typical within machine learning.
- An understanding of a range of methods of enquiry in statistics, and their suitability to specific investigations; and the ability to select and apply a range of methods to resolve problems or introduce change within a practice.
- The ability to identify, analyse, evaluate, critically reflect on and address complex problems, applying evidence-based solutions and theory-driven arguments.
- The ability to develop appropriate processes of information gathering for a given context or use; and the ability to independently validate the sources of information and evaluate and manage the information

STUDENT SUPPORT

Please contact your lecturer for subject-related support. The lecturers presenting this subject are:

- E. Cheteni – cheteni.e@belgiumcampus.ac.za

If the lecturers were unable to assist, you can also contact the cluster head for this subject:

- Ms P. Tavagwisa – privilegec.tavagwisa@belgiumcampus.ac.za

Further student support services are available via the counsellors:

- Lethlabile L. Selamolela – selamolela.l@belgiumcampus.ac.za
- Mathapelo Leshilo – leshilo.m@belgiumcampus.ac.za

ASSESSMENT PLAN			
ASSIGNMENTS/PROJECTS			
Project 1 weight:	30	Project 1 due date:	09 May 2024
Project 2 weight:	30	Project 2 due date:	13 May 2024
TESTS			
Test weight:	40	Test date:	16 May 2024

STUDENT RESOURCES Which resources will be used during this module?	
PRESCRIBED MATERIAL	
Textbook 1	
Müller, A. C & Guido, S. (2017) *Introduction to Machine Learning with Python*. O'Reilly Media, Inc.	
Location (Library / URL / PDF)	*Introduction to Machine Learning with Python.* A Guide for Data Scientists
Textbook 2	
Godsey, B. (2017) *Think Like a Data Scientist.* Manning Publication Co.	
Location (Library / URL / PDF)	Think Like a Data Scientist
Reference Textbook	
James, G., Witten, D., Hastie, R. T., & Taylor, J (2021) *An Introduction to Statistical Learning with Application in Python*	
Location (Library/URL/PDF)	
RECOMMENDED READING	
Arthur Turrell, A., Monticone, P., Mele, A., Alshannikov, I., Durrani, U., Akyol, Z., Huang, Y., Chiu, W. (2022) *Python for Data Science Handbook.* Py4DS Community.	
Location (Library / URL / PDF)	Python for Data Science Handbook
STUDENT MATERIAL	
Item	**Location**
Content on Moodle	The relevant Moodle course
Python Notebook files	Distributed to students via Moodle
Exercises / Activities	Dispersed throughout the course on Moodle. Some quizzes to be hosted on AssessmentQ.
TECHNOLOGY (HARDWARE OR SOFTWARE) REQUIRED	
Software/Hardware	**Responsible lecturer, availability on campus, recommended versions**
Microsoft Office	

LESSON PLAN OUTLINE	
Date	**Specific outcomes (SO) to be covered / Class Activity / Assessment**
29-04-2024	SO1: Introduction & Project Setup
30-04-2024	SO2: Data Science Workflow
01-05-2024	Worker's Day
02-05-2024	SO3: House Price Prediction Using Regression Algorithms
03-05-2024	SO4: Q & A session 1
06-05-2024	SO5: Customer Churning using Classification Algorithms
07-05-2024	SO6: Customer Segmentation with KMeans
08-05-2024	SO7: Stock Price Prediction Using Time Series
09-05-2024	SO8: Sentiment Analysis using nltk ***Project 1 Due***
10-05-2024	Graduation Day
13-05-2024	SO9: Q & A Session 2 ***Project 2 Due***
14-05-2024	SO10: Project 1 Presentations
15-05-2024	SO11: Project 2 Presentations
16-05-2024	**Summative Test**
17-05-2024	Career Day

OUTCOME BREAKDOWN

Specific Outcome 1: Introduction & Project Setup

- Project Setup:
  - Installation of Python & Anaconda Navigator
  - Installation of GitHub Desktop
  - Setting up a project on a local Repository linked to a GitHub repository
- Foundational skills

Specific Outcome 2: Data Science Workflow

- Business Understanding,
- Data Understanding,
- Data Preparation,
- Modelling,
- Evaluation,
- Deployment

Specific Outcome 3: House Price Prediction Using Regression Algorithms

- Prepare data – using functions for reproducibility
- Predict House Prices using Linear & Non-Linear Regression Techniques
- Evaluation *(MAE)*
- Parameter Tuning: Cross-Validation
- Reporting: Use ChatGPT to create WebApp to deploy locally

Specific Outcome 4: Q & A session 1

- Tasks allocation – follow CRISP-DM workflow (*6 phases*)
- Providing guidelines – project scope, model performance, cost optimisation
- Discuss the way forward

Specific Outcome 5: Customer churning using Classification Algorithms

- Exploratory Data Analysis
- Modelling & Evaluation (*Confusion Matrix)*
- Parameter Tuning (*Imbalanced classes)*
- Reporting: Use ChatGPT to create WebApp to deploy locally

Specific Outcome 6: Customer Segmentation with KMeans

- Exploratory Data Analysis
- Dimensionality Reduction - PCA
- Modeling & Evaluation
- Reporting: financial impact

Specific Outcome 7: Stock Price Prediction Using Time Series

- Work with APIs (*yfinance*)
- SQL databases (*SQLite*) – reproducibility matters
- Exploratory Data Analysis
    - moving averages,
    - stationarity,
    - daily returns
- use ML to predict the stock price
- Reporting: Use ChatGPT to create WebApp to deploy locally

Specific Outcome 8: Sentiment Analysis using nltk

- Data Preparation
    - Tokenization,
    - Stematization
- Modelling & Evaluation
- Reporting – Classify Reviews