COP528 AI and Applied Machine Learning Coursework

In this coursework you will apply what you have learnt in this module to solve real-world problems, both with classical machine learning and popular deep learning methods. The coursework has two tasks, the first task is about designing a pipeline (i.e., a standard procedure on data preliminary analysis, selection of ML methods, training and test data splitting, metrics selection and evaluation) to use ML to solve a predefined task, and the second task is about using a Convolutional Neural Network (CNN) for image analysis. **You can pick either of the tasks to work and write your report.**

Task 1: Machine Learning Pipeline

Design a pipeline, an evaluation strategy, and a set of experiments to determine the best parameters and machine learning algorithm, based on the results of empirical evaluations derived from a dataset (for achieving this, you could compare different algorithms if needed). Data used for this task:

A business manager of a consumer credit card portfolio is facing the problem of customer attrition. They want to analyse the data to find out the reason behind this and leverage the same to predict customers who are likely to drop off.

Multiple attributes have been provided about the customers information along with demographics.

Features:

- Client number. Unique identifier for the customer holding the account.
- Demographic variable - Customer's Age in Years.
- Demographic variable – Gender M=Male, F=Female.
- Demographic variable - Number of dependents.
- Demographic variable - Educational Qualification of the account holder (example: high school, college graduate, etc.).
- Demographic variable – Marital_Status: Married, Single, Divorced, Unknown.
- Demographic variable - Annual Income Category of the account holder (< $40K, $40K - 60K, $60K - $80K, $80K-$120K, > $120K, Unknown).
- Product Variable - Type of Card (Blue, Silver, Gold, Platinum).
- Month on book - Period of relationship with bank.
- Total relationship count: Total no. of products held by the customer.
- Months inactive: No. of months inactive in the last 12 months
- No. of Contacts in the last 12 months.
- Credit Limit on the Credit Card.
- Total Revolving Balance on the Credit Card.
- Open to Buy Credit Line (Average of last 12 months).
- Change in Transaction Amount (Q4 over Q1).
- Total Transaction Amount (Last 12 months).
- Total Transaction Count (Last 12 months).
- Change in Transaction Count (Q4 over Q1)
- Average Card Utilization Ratio.

- Attrition Flag. Internal event (customer activity) variable - if the account is closed then 1 else 0.

Inspiration

Predict whether a customer is likely to cancel his account.

Task 2: Deep Learning for image classification

You will be provided with an image dataset, where each image contains meaningful objects, e.g., parachute, oil box and truck etc. You could either use an existing CNN network architecture or design your own CNN network for classifying test images into the predefined classes.

Task 1 and 2 submission guidelines

Submission: The code, report and reflection must be submitted on 11:59AM Friday [17th March 2023] of this module.

- Project Report: PDF, up to 6 pages (IEEE double column format, template will be provided) on Learn.
- Software: Jupyter Notebook. Code needs to be submitted as attachments.

Assessment Criteria: The project will be marked based on the code quality and the report (as described below). The marking criteria is given as below.

Code:

- **Code Quality (10%):** You need to submit clean, structured and well commented code so that the instructor could run your code and get the evaluation results.

Report:

- **Introduction (10%)**: Describe the problem you are working on, application background, the machine learning task(s), and an overview of your results.
- **Data and preliminary analysis (10%)**: Briefly describe the dataset. You could use some visualisation or statistical methods to make assumptions that will influence your design.
- **Methods (20%)**: Present your machine learning approach for solving the task. The proposed approaches should be evidenced by a working piece of code of your software. You should demonstrate that you have applied ideas and skills as well as theoretical understanding of your selected methods. It may be helpful to include figures, diagrams, or tables to describe your method.
- **Experiments (20%)**: Present and discuss the experiments that you performed for the task. You may show what things you tried, what hyperparameters or architectural choices you tested, model training and evaluation strategies, what is your best model, impact of various components of your system. Justify the methods/parameters when applicable. Quantitatively evaluate and/or

compare your results, e.g. performance metrics, statistical tests, learning curve and plots. You should include graphs, tables, or other figures to illustrate your experimental results.

- **Reflection (20%)** Summarise your key results, what problems did you encounter? What are the good findings and what could make your performance better if you try it again?
- **Writing/Formatting/ Referencing (10%)** Is the report well-structured, clearly written and nicely formatted? Is the technical content presented at the right level, concise, and focused? Is the code well formatted with helpful commenting when necessary?

Further Information:

- You are recommended to use Python scikit-learn, TensorFlow and Keras for implementation. You can use data pre-processing code blocks that are available online (to avoid plagiarism, ensure that you provide the source and acknowledge the author of the code you have used. You must add this information in a comment accompanying the code fragment that you borrowed. If you have adapted the code, state "Code adapted from: [provide the source]". If you have copied the code and have made no changes to it, state "Code copied from: [https://uark.libguides.com/CSCE/CitingCode]").
- It's ok if your results are not "good". What matters is that you demonstrate your knowledge and effort made to gain a good understanding and practical skills related to machine learning in detail through the project.
- The report structure and contents are indicative. Components which are relevant to your project should be demonstrated. Some questions may be standard for your project and only a brief mention is enough. You do not need to address all of them in full detail.
- You may consult any textbooks, online resources, or publicly available implementations for ideas and code that you may adapt into your strategy or algorithm. You need to clearly cite your sources in your code and your writeup. You should not use another students' code for the class for your project.
- IT contact: science.it@lboro.ac.uk.