Machine Learning 2 – BIA 5402 Midterm Project

Schedule

- **Project time:** Entire class time.
- **Wrap up and submission:** 10 minutes.
- **Presentation:** There will be no presentation.

Overview

The midterm consists of 7 projects, each covering one key topic: Perceptron, Multilayer ANN, Regression Forecasting, Time Series Forecasting, Association Rules, ANN predictor.

Each team will:

- Use the attached Gym Members Exercise Dataset
- Build and evaluate models using Python libraries

Deliverables:

1. A PDF report, APA format, that doesn't exceeding 2000 words covering:
    - Problem definition
    - Data preprocessing
    - Model building and evaluation
    - Interpretation of results
    - Answers to reflection questions
    - Business value of the developed model. You may include plots, figures, etc. as needed.
    - Individual contribution table.

2. PowerPoint Presentation summarizing the key points (no presentation is required).
    - Focus on Business problem, results, and solution value.

3. Coding file in ipynb format that demonstrates model working on external input.

Reflection Questions (for each project)

1. What are the key features impacting the model?
2. How effective is the chosen method for the problem?
3. What challenges or limitations did you face?
4. How could this model be improved or applied practically?

About Dataset

This dataset provides a detailed overview of gym members' exercise routines, physical attributes, and fitness metrics. It contains 973 samples of gym data, including key performance indicators such as heart rate, calories burned, and workout duration. Each entry also includes demographic data and experience levels, allowing for comprehensive analysis of fitness patterns, athlete progression, and health trends.

Key Features:

- Age: Age of the gym member.
- Gender: Gender of the gym member (Male or Female).
- Weight (kg): Member's weight in kilograms.
- Height (m): Member's height in meters.
- Max_BPM: Maximum heart rate (beats per minute) during workout sessions.
- Avg_BPM: Average heart rate during workout sessions.
- Resting_BPM: Heart rate at rest before workout.
- Session_Duration (hours): Duration of each workout session in hours.
- Calories_Burned: Total calories burned during each session.
- Workout_Type: Type of workout performed (e.g., Cardio, Strength, Yoga, HIIT).
- Fat_Percentage: Body fat percentage of the member.
- Water_Intake (liters): Daily water intake during workouts.
- Workout_Frequency (days/week): Number of workout sessions per week.
- Experience_Level: Level of experience, from beginner (1) to expert (3).
- BMI: Body Mass Index, calculated from height and weight.

Grading Criteria (Total 100 Points)

Criteria	Points	Description
Data Preprocessing	**15**	**Proper handling of missing values, encoding, scaling**
Model Implementation	**20**	**Correct use of libraries, model training & tuning**
Evaluation & Metrics	**15**	**Appropriate performance measures and interpretation**
External Input Testing	**10**	**Demonstrated model prediction on new input**
Report Quality (HTML)	**15**	**Clarity, completeness, structure, and reflection**
PPT	**15**	**Clear information, and visuals**
Business value	**10**	**Ability to answer questions and explain concepts**

Projects and Requirements

Group 1: ANN – Predict Membership Retention

1. Build a neural network to classify if a member will renew their membership (Yes/No)
2. Preprocess data (encoding, scaling)
3. Use at least one hidden layer
4. Evaluate using accuracy, precision, recall, confusion matrix

Accept new input for prediction

Group 2: Multilayer ANN – Classify Feedback Score (High/Low)

1. Define feedback levels (above/below median)
2. Build ANN with at least one hidden layer
3. Use dropout/regularization
4. Show training/validation graphs
5. Accept new input for classification

Group 3: Regression-Based Forecasting – Predict Calories Burned

1. Linear regression predicting calories burned from duration, BMI, age, visits
2. Check assumptions, residuals
3. Evaluate with R^2, RMSE, MAE
4. Forecast calories burned for new input

Group 4: Time Series Forecasting – Weekly Visits Prediction

1. Use weekly visits data (20 weeks)
2. Model future visits for next 4 weeks
3. Visualize predictions
4. Forecast from new time series input

Group 5: Association Rules Mining – Member Behavior Patterns

1. Find association rules among categorical variables (Goal, Membership Type, Retention, Feedback)
2. Set support, confidence, lift thresholds
3. Present and interpret top 5 rules

Group 6: ANN Classification – Predict Member Goal Category

1. Classify member goal (Weight Loss, Fitness, Strength) with ANN
2. Preprocess data, train/test split
3. Evaluate classification metrics
4. Accept new input for goal prediction

Group 7: Experience Level Prediction

1. Build a neural network model to predict a gym member's Experience_Level (1 = Beginner, 2 = Intermediate, 3 = Expert).
2. Preprocess data (encoding, scaling)
3. Use at least one hidden layer
4. Evaluate using accuracy, precision, recall, confusion matrix
5. Accept new input for prediction.

Group 8: Workout & Health Behavior Patterns Objective

1. Discover hidden relationships between workout habits and health indicators (e.g. but not limited to, Workout_Type, BMI Category, Experience_Level, Water_Intake Level).
2. Set support, confidence, lift thresholds
3. Generate and rank association rules
4. Present and interpret the top 5 strongest rules

Group 9: Time Series Forecasting – Calories Burned Trend

1. Forecast future calories burned based on historical workout patterns.
2. Split the data into training (80%) and validation (20%) and compare at least two functions (linear, exponential, polynomial) and report the one with lower Standard Error
3. Build a regression model that can used to forecast calories burned for the next periods.
4. Visualize historical data and forecast.

Group 10: ANN Classification – Workout Intensity Level

1. Classify workout sessions into Low / Medium / High intensity using ANN.
2. Define workout intensity using physiological thresholds (BPM, calories, duration)
3. Build ANN with at least one hidden layer
4. Accept new workout input and classify intensity