Marking Scheme (100 Marks Total)

- **Problem Understanding & Assumptions (10 Marks):**
  Clarity in framing the chosen problem, dataset mapping, and assumptions.

- **Data Preprocessing & Feature Engineering (20 Marks):**
  Handling missing values, encoding categorical features, scaling, and creativity in feature engineering.

- **Model Selection & Justification (15 Marks):**
  Logical choice of algorithms with justification (why clustering? why regression? etc.).

- **Implementation & Code Quality (15 Marks):**
  Clean, well-structured, and commented code. Reproducible workflow.

- **Evaluation & Metrics (15 Marks):**
  Correct application of evaluation metrics (e.g., Silhouette for clustering, RMSE for regression, Accuracy/F1 for classification).

- **Visualization & Interpretation (10 Marks):**
  Quality of charts/plots to support conclusions (heatmaps, scatter plots, cluster maps, etc.).

- **Insights & Recommendations (10 Marks):**
  Actionable insights tied back to the SNU context (canteen, wellness, clubs, transport, etc.).

- **Originality & Presentation (5 Marks):**
  Creativity in framing results, uniqueness of approach, and professional report writing.

Evaluation Metrics by Problem Type

- **Clustering:** Silhouette Score, Davies–Bouldin Index, visualization of clusters.

- **Classification:** Accuracy, F1-score, Confusion Matrix.

- **Regression:** RMSE, MAE, R^2 Score.

- **Recommendation:** Precision@K, Recall@K, interpretability of recommendations.

General Instructions

1. **Dataset Source:**
    - You will collect the dataset yourselves through the Google Form provided.
    - Each student must respond truthfully. This ensures a real, authentic dataset.

2. **Deliverables:**
    - **Jupyter Notebook (.ipynb):** Must contain all preprocessing, code, visualizations, and explanation.
    - **Report (PDF):** A concise summary (2–4 pages) including problem statement, methods, results, and insights.
    - **GitHub Link:** Upload your code to GitHub and share the repository link in your submission.
    -
    - Short presentation video

3. **Submission Method:**
    - Upload all deliverables (Notebook, Report, GitHub link) into a google drive and share the link with me (bidyut.s@snuniv.ac.in) and also print the hard copy of the Report and submit it to me before the deadline. For a team only one copy is enough

4. **Deadline:**
    - Submissions close on **10th September, 2025 at 11:59 PM IST**. Late submissions will not be graded.

📘 Sister Nivedita University – Machine Learning Challenge 2025

Prepared by: **Assistant Professor Bidyut Saha, School of Engineering & Technology, SNU**

A. Clustering Challenges (Unsupervised)

1. The Friendship Blueprint of SNU
At SNU, club memberships and hobbies shape how students bond. The Student Union wants to optimize seating arrangements in common rooms to strengthen friendships.
📊 **Columns:** `club_top1`, `club_top2`, `teamwork_preference`, `hobby_top1`, `hobby_top2`
🧩 **Task:** Cluster students into "friendship groups" based on shared hobbies and club interests.
🏆 **Impact:** Helps SNU plan club activities and student hangouts more effectively.

2. Learning Archetypes at SNU
Professors at SNU notice that students learn differently – some read books, some binge series, others prefer coding. They want to design adaptive teaching strategies.
📊 **Columns:** `books_read_past_year`, `reads_books`, `book_genre_top1`, `screen_time_movies_series_hours_per_week`, `binge_freq_per_week`
🧩 **Task:** Cluster students into learning archetypes.
🏆 **Impact:** Enables personalized mentoring approaches.

3. Digital DNA of SNU Students
The SNU IT Cell wants to build a new student app. To design features, they need to understand how students behave digitally.
📊 **Columns:** `gaming_platform_top1`, `social_platform_top1`, `daily_social_media_minutes`, `ott_top1`, `content_creation_freq`
🧩 **Task:** Cluster students into digital behavior groups.
🏆 **Impact:** Helps design apps tailored to SNU student life.

4. Wellness Personas of SNU
The Health Club wants to launch wellness programs but doesn't know the lifestyle clusters at SNU.
📊 **Columns:** `eating_out_per_week`, `food_budget_per_meal_inr`, `sweet_tooth_level`, `weekly_hobby_hours`
🧩 **Task:** Cluster students into lifestyle personas (e.g., "health-conscious," "fast-food lovers").
🏆 **Impact:** Enables targeted health campaigns.

5. OTT Audience Map for SNU Film Fest
The Cultural Committee is planning the annual Film Fest. They want to know the student audience segments.
📊 **Columns:** `movie_genre_top1`, `series_genre_top1`, `ott_top1`, `content_lang_top1`
🧩 **Task:** Cluster students into viewing preference groups.
🏆 **Impact:** Helps plan screenings and OTT tie-ups.

6. Cultural-Fest Gamer-Music Matchmaker
For the cultural fest, SNU wants to pair gaming tournaments with music concerts.
📊 **Columns:** `game_genre_top1`, `game_genre_top2`, `music_genre_top1`, `music_genre_top2`, `listening_hours_per_day`
🧩 **Task:** Cluster students to find gamer–music overlaps.
🏆 **Impact:** Increases event participation and engagement.

🔹 B. Classification Challenges (Supervised)

7. Group or Solo? The Project Pairing Dilemma
Faculty often struggle to form balanced project teams. Some students prefer working solo, while others thrive in groups.
📊 **Target:** `teamwork_preference`
📊 **Features:** `introversion_extraversion`, `risk_taking`, `club_top1`, `weekly_hobby_hours`
🧩 **Task:** Predict whether a student prefers solo or group projects.
🏆 **Impact:** Helps faculty assign balanced teams.

8. The Canteen Menu Optimizer

The SNU canteen manager wants to predict dietary preferences to stock food items better.

📊 **Target:** `dietary_preference`

📊 **Features:** `cuisine_top1`, `spice_tolerance`, `sweet_tooth_level`

🧩 **Task:** Predict diet type (Veg/Non-Veg/Vegan/etc.).

🏆 **Impact:** Reduces food waste, improves menu planning.

9. **Tea vs Coffee – SNU Café Wars**

The SNU Café is debating whether to expand tea or coffee options.

📊 **Target:** `tea_vs_coffee`

📊 **Features:** `age`, `dietary_preference`, `daily_social_media_minutes`, `introversion_extraversion`

🧩 **Task:** Predict beverage preference.

🏆 **Impact:** Informs café business decisions.

10. **Finding SNU's Hidden Athletes**

Many students quietly play sports but aren't part of official teams. The Sports Council wants to identify them.

📊 **Target:** `hobby_top1` (sports vs non-sports)

📊 **Features:** `weekly_hobby_hours`, `teamwork_preference`, `risk_taking`

🧩 **Task:** Predict if a student is a hidden athlete.

🏆 **Impact:** Helps recruit for inter-university competitions.

11. The OTT King or Queen of SNU

Sponsors want to know which OTT platform dominates student life.

📊 **Target:** `ott_top1`

📊 **Features:** `movie_genre_top1`, `series_genre_top1`, `binge_freq_per_week`, `screen_time_movies_series_hours_per_week`

🧩 **Task:** Predict dominant OTT platform.

🏆 **Impact:** Helps secure sponsorship deals.

12. Who Will Speak Up in Class?

Faculty want to identify students who are most likely to ask questions and participate actively.

📊 **Target:** `introversion_extraversion`

📊 **Features:** `reads_books`, `risk_taking`, `club_top1`

🧩 **Task:** Predict likelihood of participation.

🏆 **Impact:** Improves classroom engagement strategies.

13. **Gamer Hunt for SNU Gaming League**

The Gaming Club needs to identify regular gamers to recruit for e-sports tournaments.

📊 **Target:** `gaming_days_per_week` (>3 = gamer)

📊 **Features:** `game_genre_top1`, `gaming_platform_top1`, `esports_viewing`

🧩 **Task:** Predict if a student is a regular gamer.

🏆 **Impact:** Builds stronger e-sports teams.

🔹 C. Regression Challenges (Supervised)

14. **The Mentor's Study Time Predictor**

Faculty want to estimate how much time each student studies daily for better mentoring.

📊 **Target:** `books_read_past_year` (proxy for study hours)

📊 **Features:** `reads_books`, `book_genre_top1`, `screen_time_movies_series_hours_per_week`

🧩 **Task:** Predict study time.

🏆 **Impact:** Enables better mentorship allocation.

15. **The Sleep Health Report**

The SNU Wellness Cell wants to predict sleep hours based on lifestyle indicators.

📊 **Target:** `sleep_hours`

📊 **Features:** `daily_social_media_minutes`, `gaming_hours_per_week`, `introversion_extraversion`

🧩 **Task:** Predict daily sleep hours.

🏆 **Impact:** Helps design awareness campaigns.

16. **SNU Transport Planner**

The Transport Committee wants to optimize bus schedules.

📊 **Target:** `commute_time` (self-reported)

📊 **Features:** `age`, `daily_social_media_minutes`, `weekly_hobby_hours`

🧩 **Task:** Predict commute time.

🏆 **Impact:** Improves transport planning.

17. **Coding Hours Forecaster for SNU Hackathons**
Hackathon organizers need to balance teams by coding effort.
📊 **Target:** `weekly_hobby_hours` (coding subset)
📊 **Features:** `hobby_top1`, `club_top1`, `reads_books`
🧩 **Task:** Predict coding hours per student.
🏆 **Impact:** Ensures fair and balanced hackathon teams.

18. **Canteen Budget Planner**

The canteen wants to know how much students typically spend per meal.

📊 **Target:** `food_budget_per_meal_inr`

📊 **Features:** `dietary_preference`, `eating_out_per_week`, `age`, `fashion_spend_per_month_inr`

🧩 **Task:** Predict meal budget.

🏆 **Impact:** Helps keep prices student-friendly.

🔹 D. Recommendation & Derived-Label Challenges

19. The SNU Study Buddy Finder
Students often struggle to find the right study partner. The Dean's Office wants a buddy recommendation system.
📊 **Columns:** `teamwork_preference`, `introversion_extraversion`, `books_read_past_year`, `club_top1`, `weekly_hobby_hours`
🧩 **Task:** Recommend compatible study buddies.
🏆 **Impact:** Improves collaborative learning outcomes.

20. Hobby Expansion for SNU Clubs

SNU's clubs want to suggest new hobbies for members based on their lifestyle.

📊 **Columns:** `hobby_top1`, `hobby_top2`, `club_top1`, `music_genre_top1`, `game_genre_top1`

🧩 **Task:** Recommend new hobbies/clubs to students.

🏆 **Impact:** Increases participation and enriches student life.

Personal Information

- `age` — int — 16–30 — numeric
- `height_cm` — int — 120–220 — numeric
- `weight_kg` — int — 30–150 — numeric

Food

- `cuisine_top1`, `cuisine_top2`, `cuisine_top3` — string — {Indian, Chinese, Thai, Italian, Mexican, Mughlai, Bengali, South Indian, Mediterranean, Japanese, Korean, Continental, Street Food, Vegan} — label encode / one-hot
- `spice_tolerance` — int — 1–5 — numeric (Likert)
- `dietary_preference` — string — {Veg, Non-Veg, Eggitarian, Vegan, Jain} — one-hot/label
- `eating_out_per_week` — int — 0–10 — numeric
- `food_budget_per_meal_inr` — int — 50–1500 — numeric
- `sweet_tooth_level` — int — 1–5 — numeric (Likert)
- `tea_vs_coffee` — string — {Tea, Coffee, Both, Neither} — one-hot/label

Movies & Series

- `movie_genre_top1`, `movie_genre_top2`, `movie_genre_top3` — string — {Action, Drama, Comedy, Romance, Sci-Fi, Horror, Thriller, Animation, Documentary, Biopic} — one-hot/label

- `series_genre_top1`, `series_genre_top2`, `series_genre_top3` — string — {Crime, Sitcom, Fantasy, Historical, Teen, K-Drama, Anime} — one-hot/label

- `content_lang_top1`, `content_lang_top2`, `content_lang_top3` — string — {English, Hindi, Bengali, Tamil, Telugu, Kannada, Malayalam, Marathi, Other} — one-hot/label

- `ott_top1`, `ott_top2`, `ott_top3` — string — {Netflix, Prime Video, Disney+ Hotstar, JioCinema, YouTube, SonyLIV, ZEE5} — one-hot/label

- `binge_freq_per_week` — int — 0–7 — numeric

- `screen_time_movies_series_hours_per_week` — int — 0–40 — numeric

Games

- `gaming_days_per_week` — int — 0–7 — numeric

- `gaming_hours_per_week` — int — 0–50 — numeric

- `game_genre_top1`, `game_genre_top2`, `game_genre_top3` — string — {FPS, MOBA, RPG, Sports, Racing, Strategy, Casual, Puzzle} — one-hot/label

- `gaming_platform_top1`, `gaming_platform_top2` — string — {Mobile, PC, Console, Cloud} — one-hot/label

- `esports_viewing` — string — {Never, Sometimes, Often} — ordinal map (0/1/2) or one-hot

Social Media

- `social_platform_top1`, `social_platform_top2`, `social_platform_top3` — string — {Instagram, WhatsApp, YouTube, Facebook, X/Twitter, Reddit, LinkedIn, Snapchat, Telegram} — one-hot/label

- `daily_social_media_minutes` — int — 0–600 — numeric

- `primary_content_type` — string — {Memes, News, Educational, DIY/Coding, Lifestyle, Gaming, Music} — one-hot/label

- `content_creation_freq` — string — {No, Occasional, Regular} — ordinal map (0/1/2) or one-hot

Music

- `music_genre_top1`, `music_genre_top2`, `music_genre_top3` — string — {Bollywood, Classical, Indie, Pop, Rock, Hip-Hop, EDM, Lo-fi, Devotional} — one-hot/label
- `listening_hours_per_day` — float — 0–10 — numeric
- `music_lang_top1`, `music_lang_top2` — string — {English, Hindi, Bengali, Tamil, Telugu, Punjabi, Other} — one-hot/label
- `live_concerts_past_year` — int — 0–10 — numeric

Reading Habits

- `reads_books` — string — {No, Sometimes, Regularly} — ordinal (0/1/2) or one-hot
- `book_genre_top1`, `book_genre_top2`, `book_genre_top3` — string — {Fiction, Non-Fiction, Self-Help, Tech, Biography, Sci-Fi, Fantasy} — one-hot/label
- `books_read_past_year` — int — 0–50 — numeric

Shopping Preferences

- `fashion_spend_per_month_inr` — int — 0–20000 — numeric
- `shopping_mode_preference` — string — {Mostly Online, Mixed, Mostly Offline} — one-hot/ordinal
- `ethical_shopping_importance` — int — 1–5 — numeric (Likert)

Travel

- `travel_freq_per_year` — int — 0–12 — numeric
- `travel_type_top1`, `travel_type_top2`, `travel_type_top3` — string — {Trekking, Beach, City, Pilgrimage, Road Trip, Cultural} — one-hot/label
- `budget_per_trip_inr` — int — 1000–100000 — numeric

- `travel_planning_preference` — int — 1–5 — numeric (Likert; 1=Spontaneous, 5=Highly Planned)

Hobbies & Clubs

- `hobby_top1`, `hobby_top2` — string — {Coding, Photography, Dance, Music, Painting, Writing, Cricket, Football, Badminton, Gym, Yoga, Theatre, Debate, Robotics, Hackathons} — one-hot/label

- `club_top1`, `club_top2` — string — {Coding Club, Robotics Club, Cultural Club, Sports Club, Drama Club, Music Club, Literary Club, Entrepreneurship Cell} — one-hot/label

- `weekly_hobby_hours` — int — 0–40 — numeric

Personality Brief

- `introversion_extraversion` — int — 1–5 — numeric (Likert)

- `risk_taking` — int — 1–5 — numeric (Likert)

- `conscientiousness` — int — 1–5 — numeric (Likert)

- `openness_to_new_experiences` — int — 1–5 — numeric (Likert)

- `teamwork_preference` — int — 1–5 — numeric (Likert; 1=Prefer Solo, 5=Prefer Teams)