KNime Practice session (17th Jan 2026)

Dataset Link
: https://drive.google.com/file/d/1Slujv63b1AmLGUZ7XsfE6Sc807XvHmTu/view?usp=sharing

🧮 **PART A: Linear Regression Task**

Objective:

Build and evaluate **Linear Regression models** to predict the Final_Score.

Tasks:

1. Import the dataset student_performance_dataset.csv into KNIME.
2. Split data using **Partitioning Node** (e.g., 70% Train – 30% Test).
3. Apply **Linear Regression Learner Node** on training data with different configurations:
    - **Model 1:** Use all features (default settings).
    - **Model 2:** Exclude one feature (e.g., Social_Media_Hours) and observe the change.
4. Use the **Linear Regression Predictor Node** on the test data.
5. Connect **Numeric Scorer Node** to compute **RMSE** and **R^2** scores.

6. Comment in KNIME nodes:
    - “This model uses normalized input features to test impact on RMSE and R².”
    - “This model excludes Social_Media_Hours to test multicollinearity effects.”

✅ **Expected Output:**

- RMSE and R² values for all 3 models.
- Observation on which model performs best and why.

🔢 **PART B: Classification Task**

Objective:

Build **Classification models** to predict whether a student will “Pass” or “Fail”.

Target Column: Pass_Status

Tasks:

1. Convert target variable to categorical using **Category to Number** or **One-Hot Encoding** if needed.

2. Partition dataset into 70% training and 30% testing.

3. Build and compare multiple models:

   Model 1 & 2 – Logistic Regression:

   - Model 1: Default settings.
   - Model 2: Change maximum iterations or regularization parameter.

   Model 3 & 4 – Decision Tree:

   - Model 3: Default depth.
   - Model 4: Max Depth = 5, Minimum Split Size = 10.

   Model 5 & 6 – Random Forest:

   - Model 5: 50 trees.
   - Model 6: 100 trees and Maximum Depth = 8.

4. Use the **Scorer Node** to measure:
    - **Accuracy**
    - **Precision**
    - **Recall**
    - **F1-Score**

5. Comment in each model:
    - “Model 2 Logistic Regression uses increased iterations to improve convergence.”
    - “Decision Tree Model 4 controls overfitting by limiting depth.”
    - “Random Forest Model 6 increases tree count for better generalization.”

✅ **Expected Output:**

- Confusion matrices and performance metrics for all models.
- A short written note comparing the results (which model gives best accuracy and why).

* Indicates required question

Full Name *

Your answer

Registered Mail id *

Your answer

Upload Knime file *

Upload up to 5 supported files. Max 10 MB per file.

Add File

Submit Clear form