**🔍 Project Title**

**Comparative Study of Classification Algorithms on Public Datasets**

**📌 Step-by-Step Guideline (Beginning to End)**

**1. Team Planning & Dataset Selection**

* Choose dataset(s): e.g., Iris (simple), Titanic (requires preprocessing), or MNIST (image data).
* Assign roles: Data preprocessing, Model training, Performance analysis.
* Tools: Jupyter Notebook/Colab, scikit-learn, pandas, matplotlib/seaborn.

**2. Literature Review**

* Briefly study and document:
  + How each algorithm works (SVM, k-NN, Decision Tree, Logistic Regression)
  + Previous papers comparing these models
  + Dataset characteristics

**3. Data Preprocessing**

* Handle missing values, encode categorical data, normalize/standardize if necessary.
* Apply different **feature engineering strategies**: PCA, One-hot encoding, scaling, etc.

**4. Model Implementation**

* Train each algorithm using train\_test\_split.
* Use proper parameter tuning (GridSearchCV or manually).
* Evaluate using:
  + Accuracy, Precision, Recall, F1-Score
  + Confusion matrix
  + ROC-AUC curve (for binary classification)

**5. Comparison & Analysis**

* Tabulate results.
* Discuss:
  + Trade-offs (speed, accuracy, interpretability)
  + Performance under different preprocessing
  + Which algorithm performs better and why

**6. Write the Research Paper**

Sections:

* **Abstract**
* **Introduction**
* **Literature Review**
* **Methodology**
* **Experiments**
* **Results & Discussion**
* **Conclusion**
* **References**

**7. Presentation (Optional)**

* Use charts to show performance comparison.
* Highlight key findings and insights.

**🗓️ Suggested 7-Day Routine for 3 Members**

| **Day** | **Tasks** | **Assigned To** |
| --- | --- | --- |
| **Day 1** | Finalize dataset, divide responsibilities, and review related literature. | All members |
| **Day 2** | Preprocess dataset and try 1-2 feature engineering strategies. | Member 1 |
| **Day 3** | Train Logistic Regression & Decision Tree models. Document results. | Member 2 |
| **Day 4** | Train SVM & k-NN models. Document results. | Member 3 |
| **Day 5** | Perform comparative analysis of all models. Create visuals. | All members |
| **Day 6** | Draft and refine research paper sections. | All members |
| **Day 7** | Final proofreading, APA formatting, citations, and submission prep. | All members |

**✅ Tips**

* Use classification\_report() and cross\_val\_score() for robust evaluation.
* Use Git or Google Drive to collaborate.
* Optional: Try multiple datasets for stronger insights.

**✅ What Makes a Comparative Study Worth Publishing?**

1. **Clear Research Question**  
   Instead of just "Which algorithm performs better?" ask:
   * *"How does feature engineering affect classical classifiers across datasets with different characteristics?"*
   * *"Which classification algorithm performs best with minimal preprocessing?"*
2. **Novelty or Depth**
   * Test on **multiple datasets** (structured & unstructured, balanced & imbalanced).
   * Include **runtime/complexity benchmarks**.
   * Add **visual comparison** (e.g., decision boundaries, ROC curves).
   * Try different **cross-validation techniques** or **imbalanced data handling** (SMOTE, undersampling, etc.).
3. **Insightful Discussion**  
   Go beyond accuracy numbers. Include:
   * Why did a certain model perform better?
   * Where did a model fail? Any pattern?
   * How easy is each model to tune/interpret?
4. **Reproducibility**
   * Share code (GitHub/Colab).
   * Share dataset links.
   * Document parameters, environment used, libraries, etc.
5. **Proper Academic Format**
   * Follow IEEE, ACM, or APA depending on where you’re submitting.
   * Include structured abstract, keywords, methodology, results, and references.

**🏛️ Where Can You Publish?**

1. **Student & Undergraduate Journals**
   * *International Journal of Undergraduate Research and Creative Activities*
   * *Undergraduate Research Journals at your university*
   * *IJARIIT*, *IJRASET*, *IJERT* (easier acceptance, often fee-based)
2. **Conferences**
   * Student Research Conferences (check your university or country)
   * International virtual conferences on AI/ML
   * Poster sessions at tech fests or symposiums
3. **arXiv (Preprint Repository)**
   * Great for early visibility.
   * No peer-review, but widely respected.

**📈 How to Increase Acceptance Chances**

* Use **at least 2 datasets** with varied characteristics.
* Show **impact of different preprocessing techniques**.
* Include **feature importance** or **SHAP values** if possible.
* Cite recent papers doing similar work (Google Scholar helps).