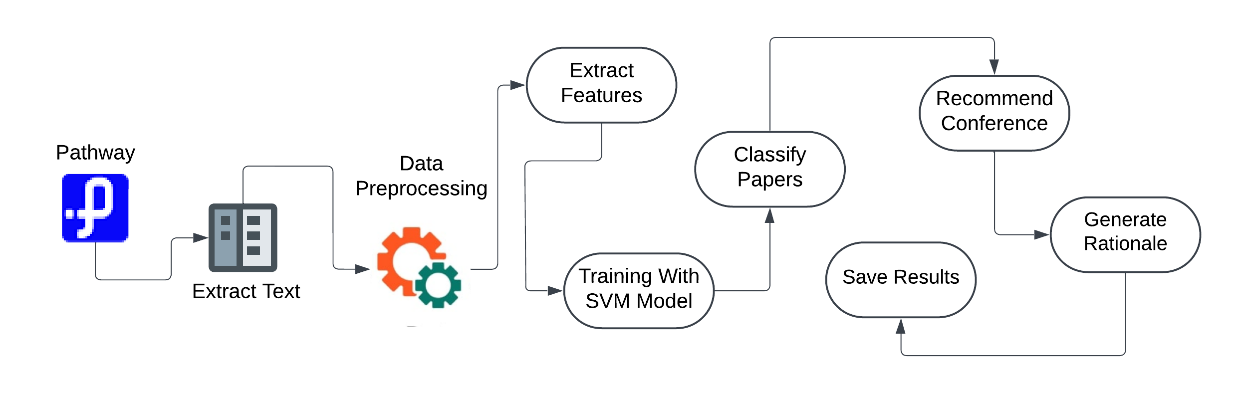
***Kharagpur Data Science Hackathon 2025***

**Report on Research Paper Evaluation and Conference Recommendation**

**Objective :**To automate the classification of research papers into "Publishable" or "Non-Publishable" categories and recommend suitable conferences for publishable papers using ML Techniques.

***Flow of Our Model :***

****

**What We Did :**

1. **Data Loading** : Used Pathway's Google Drive connector to dynamically fetch PDFs from a shared folder provided in the Google Drive.
2. **Text Extraction** : Extracted text from PDFs using PyPDF2 Pyton Library.
3. **Preprocessing** : Cleaned text by removing special characters, digits, and stopwords using NLTK and also Converted the text to lowercase for uniformity.
4. **Feature Extraction** : Computed features such as word count, lexical diversity, and reference count and also used TfidfVectorizer for vectorizing textual data.
5. **Classification** : Trained an SVM model with hyperparameter tuning using GridSearchCV to classify papers.
6. **Conference Recommendation** : Built another SVM model to recommend conferences for publishable papers and Encoded conference labels using LabelEncoder.
7. **Output** : Generated a CSV file containing paper ID, filename, publishability, recommended conference, confidence, and rationale.

**Key Tools :**

* **Pathway**: For data integration and processing.
* **PyPDF2**: For extracting text from PDFs.
* **Scikit-learn**: For machine learning and feature extraction.
* **NLTK**: For text preprocessing.

**Team Name :** Constellation

**Team Members :** Swayam Takkamore, Yash Dharme, Shlok Talhar, Devansh Paltewar

**College Name :** St. Vincent Pallotti College of Engineering and Technology