



**Project Title:** Book Genre Classification  
using Metadata

**Name:** Animesh Chandra Srivastava

**Roll No.:** 202401100300042

**Course:** CSE AI

**Section:** A

**Assessment:** Mid-Semester Examination 2  
(MSE2)

---

## Introduction

The aim of this project is to classify books into genres based on various metadata attributes such as author popularity, book length, and the number of keywords associated with the book. With the rising amount of digital content, an automated system for genre classification helps in effective cataloging, improving recommendation systems, and enhancing user experience. The problem was tackled using Python, with model training and evaluation done on Google Colab.

---

## Methodology

1. **Dataset:** The dataset contained columns including author\_popularity, book\_length, num\_keywords, and the target column genre.

2. **Preprocessing:**

- Label encoding was used to convert genre labels into numerical form.
- StandardScaler was applied to normalize the feature values.
- 

3. **Modeling:**

- A RandomForestClassifier was trained using 80% of the data.
- The remaining 20% was used for testing.

#### **4. Evaluation:**

- Classification report provided precision, recall, and F1-score.
  - Confusion matrix and feature importance were visualized using Seaborn.
- 

#### **Code**

# Full code for preprocessing, modeling, and visualization here.

# (Refer to the provided final code in the main Python file)

---

**Output/Result**

*Classification Report (sample):*

	precision	recall	f1-score	support
fantasy	0.29	0.50	0.36	4
fiction	1.00	1.00	1.00	1
mystery	0.67	0.60	0.63	10
non-fiction	0.67	0.40	0.50	5
accuracy			0.55	20
macro avg	0.65	0.62	0.62	20
weighted avg	0.61	0.55	0.56	20

*Visual Outputs:*

- 1. Genre Distribution Bar Graph
  - 2. Feature Importance Bar Plot
  - 3. Confusion Matrix Heatmap
-

## References/Credits

- Scikit-learn: <https://scikit-learn.org>
  - Pandas Documentation: <https://pandas.pydata.org>
  - Seaborn Documentation:  
<https://seaborn.pydata.org>
  - Dataset: Provided as book\_genres.csv
  - Google Colab for executing and testing code
-