

School of Computer Science Engineering and Technology  
Assignment-09

<b>Course-</b> B.Tech	<b>Type-</b> Core
<b>Code-</b> 23CS106	<b>Course Name-</b> Artificial Intelligence & Machine Learning
<b>Year-</b> 2024-2025	<b>Semester-</b> Even, <b>Instructor:</b> Prof. E.L.N. Kiran
<b>Date-</b> 30-01-2024	<b>Batch-</b> AIML-A,B

## 1 Implement Random Forest Ensemble Model using Lending Data

### 1 – Import the required Python, Pandas, Matplotlib, Seaborn packages

Problem: Lending Club connects people who need money (borrowers) with people who have money (investors). We try to create a model to predict the risk of lending money to someone given a wide range of credit related data. We will use lending data from 2007-2010 and **predict whether or not the borrower would be able to pay back their loan in full or not.**

- Features: `credit.policy:`, `int.rate:`, `installment:`, `log.annual.inc:`, `dti:`, `fico:`, `days.with.cr.line:`, `revol.bal:`, `revol.util:`
- 13 features are taken into account to decide whether to lend the money to the borrower or not.

### 2 –Perform the Data and Exploratory Analysis

1. Load the classified dataset into a dataframe using `pandas`
2. Check the data types of each feature(column) in the dataset.
3. Generate a summary of the dataset for `min`, `max`, `stddev`, `quartile vales` for 25%,50%,75%,90%,
4. List the names of columns/features in the dataset
5. Display the count of credit approvals done by the company
6. Generate a lmpot to verify if trend differed between `not_fully_paid` and `credit_policy`.

### 3 –Categorize the features and fit the data

you will categorize the data using `get_dummies` from `sklearn` package on `purpose` feature to understand what purpose borrower is lending money from company.

#### 4 – Model training and Fit the data to Model

1. Split the data generated from list created as **X**, **Y** is distributed using *train\_test\_split* function as *X\_train*, *Y\_train*, *X\_test*, *Y\_test*
2. Apply the RF Classifier model of `sklearn.ensemble` import `RandomForestClassifier` package
3. Fit the data to the Classifier Model using `fit`.

#### 5 – Evaluate the Classification Quality

1. Generate the confusion matrix to estimate the correction among features
2. Generate the classification report using `classification_report`