SYMBIOSIS
INSTITUTE OF TECHNOLOGY, NAGPUR
॥वसुधैव कुटुम्बकम्॥

DS MINI PROJECT

Analysis and Forecasting of Credit by Scheduled Commercial Banks in India (2010-2023)

Asita Ganatra
22070521003
Sem: VII
Under the guidance of
Dr./Prof. Piyush Chauhan

Symbiosis Institute of Technology, Nagpur Campus

Problem Definition

- Need for analyzing credit data of Indian Banks (2010-2023)
- Objective: Build a Streamlit-based interactive tool for:
  - Data exploration
  - Prediction (Loan/Accounts)
  - Classification and Clustering
  - Deep Learning-based pattern recognition

Problem Definition

	year	region	state_name	state_code	district_name	district_code	population_group	bank_group	occupation_group	occupation_sub_group	no_of_accounts	credit_limit	amount_outstanding
0	2014	Northern Region	Rajasthan	8	Barmer	90	Semi-Urban	Public Sector Banks	Personal Loans	Vehicles	966	39.09	28.03
1	2020	Central Region	Madhya Pradesh	23	Jhabua	412	Rural	Public Sector Banks	Personal Loans	Housing	2189	61.08	46.33
2	2012	Central Region	Madhya Pradesh	23	Shahdol	429	Rural	Public Sector Banks	Personal Loans	Education	95	2.02	1.77
3	2011	Southern Region	Telangana	36	Hyderabad	507	Urban	Public Sector Banks	Industry	Food Manufacturing & Processing	29	92.42	78.36
4	2016	Southern Region	Tamil Nadu	33	Villupuram	596	Rural	Public Sector Banks	Industry	Petroleum, Coal Products & Nuclea	10	0.21	0.18
5	2020	Central Region	Chhattisgarh	22	Korba	383	Semi-Urban	Public Sector Banks	Professional And Otl	Recreation Services	1	0.01	0.01
6	2022	Central Region	Uttar Pradesh	9	Kanpur Nagar	157	Semi-Urban	Public Sector Banks	Industry	Gems And Jewellery	1	0.02	0.02
7	2022	Northern Region	Rajasthan	8	Rajsamand	112	Semi-Urban	Private Sector Banks	Personal Loans	Education	11	0.55	0.41
8	2014	Southern Region	Tamil Nadu	33	Tiruppur	634	Semi-Urban	Public Sector Banks	All Others	All Others	410	9.79	9.09
9	2015	Northern Region	Rajasthan	8	Chittorgarh	95	Semi-Urban	Regional Rural Banks	Industry	Construction	15	0.53	0.38
10	2014	Central Region	Uttar Pradesh	9	Etah	138	Urban	Regional Rural Banks	Personal Loans	Housing	1	0.08	0.04
11	2021	Central Region	Madhya Pradesh	23	Ujjain	435	Rural	Regional Rural Banks	Industry	Chemicals & Chemical Products	6	0.07	0.06
12	2019	Western Region	Gujarat	24	Panch Mahals	454	Rural	Regional Rural Banks	Professional And Otl	Tourism, Hotel & Restaurants	15	0.1	0.07
13	2020	Eastern Region	Bihar	10	Aurangabad	189	Urban	Foreign Banks	Industry	Other Industries	1	0.02	0.02
14	2022	Eastern Region	Odisha	21	Cuttack	350	Rural	Private Sector Banks	Industry	Basic Metals & Metal Products	12	1.51	1.28
15	2015	North Eastern Regi	Arunachal Pradesh	12	Upper Subansir	241	Semi-Urban	Public Sector Banks	Industry	Mining & Quarrying	8	0.23	0.15
16	2017	Central Region	Uttar Pradesh	9	Rae Bareli	175	Urban	Public Sector Banks	Industry	Beverage & Tobacco	11	1.36	0.81
17	2017	Northern Region	Delhi	7	New Delhi	79	Rural	Public Sector Banks	Professional And Otl	It And Telecommunications	31	0.42	0.29
18	2021	Eastern Region	Odisha	21	Dhenkanal	352	Rural	Small Finance Banks	Agriculture	Direct Finance	8111	23.38	14.21

Data Collection

Dataset Used: India Data Portal (Credit by Scheduled Commercial Banks 2010-2023)

Work Done:

- Imported dataset: pandas.read_csv()
- Handled missing and inconsistent values
- Performed log transformation on skewed columns
- Encoded categorical features (LabelEncoder)
- Normalized numerical features for clustering/DNN

Data Collection

Deploy

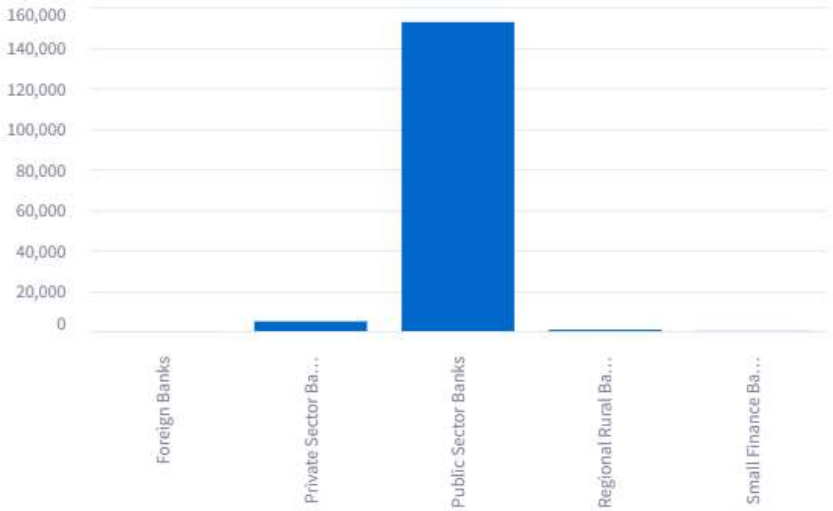
💰 Indian Bank Credit Dashboard

This dashboard analyzes scheduled commercial bank credit data.

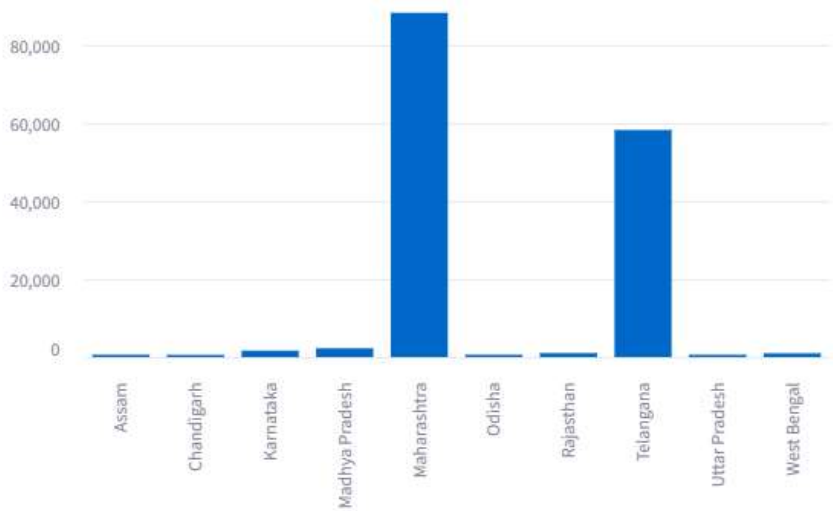
High-Level KPIs (for filtered data)

Total Amount Outstanding	Total Number of Accounts	Total Districts
₹ 1,582.18 Cr	542,768	331

Amount Outstanding by Bank Group



Top 10 States by Amount Outstanding



Exploratory Data Analysis (EDA)

What I Did:

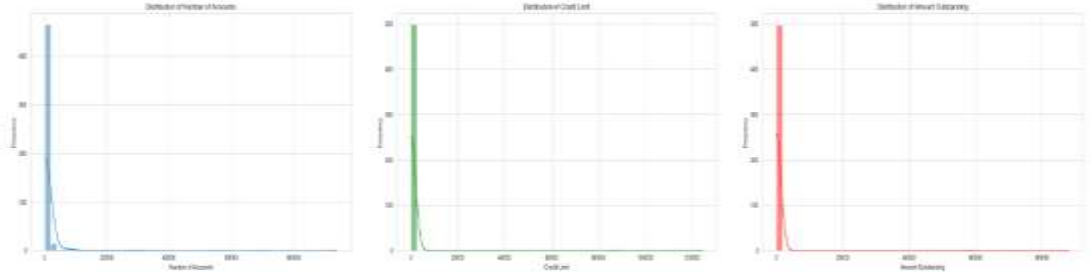
- Generated summary statistics (df.describe())
- Visualized trends:
  - Year-wise loan growth
  - Top states by amount outstanding
  - Credit distribution by bank group
- Used Matplotlib and Plotly for interactive charts
- Derived KPIs for dashboard

Exploratory Data Analysis (EDA)

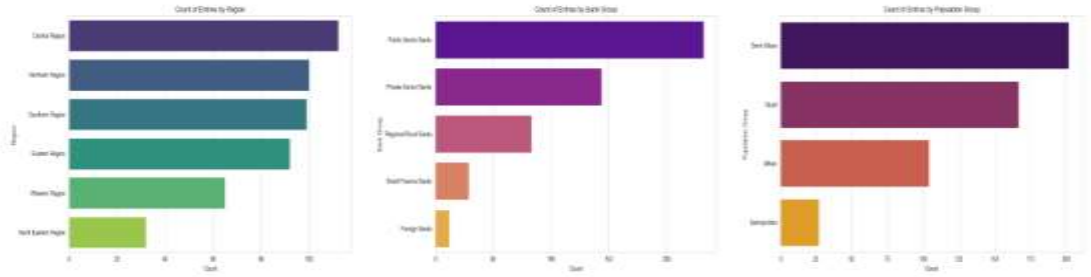
Deplo

Deploy

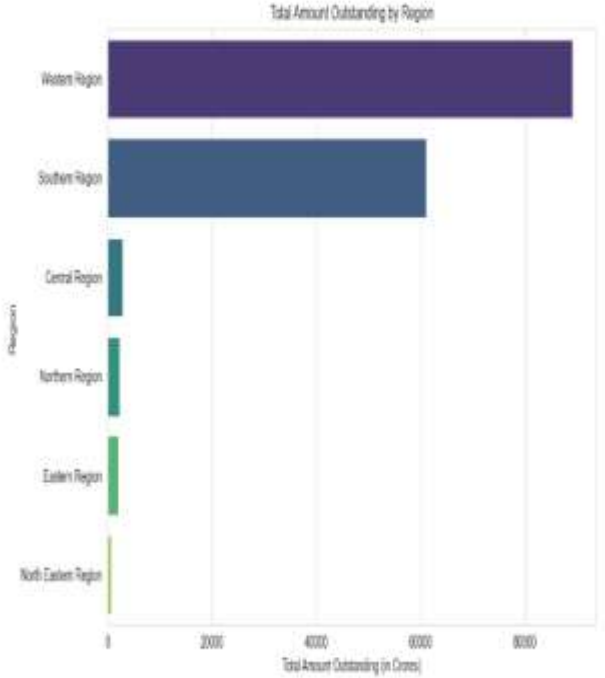
Univariate Analysis: Numerical Distributions

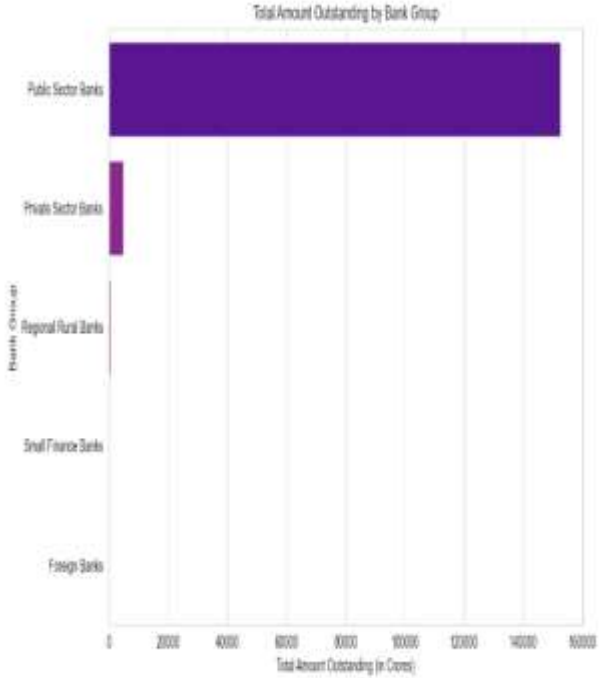


Univariate Analysis: Categorical Counts



Bivariate Analysis: Relationships and Trend





Predictive Modeling (Regression)

Goal: Forecast Loan Amount and Number of Accounts

Models Implemented:

1. **Linear Regression**
2. **Random Forest Regressor**

Work Done:

- Split data (train_test_split)
- Trained, evaluated using R2 score
- Saved models with .joblib

Results:

Prediction	Linear Regression (R2 score)	Random Forest (R2 score)
Amount Outstanding	0.6850	0.6820
Number of Accounts	0.7150	0.6894

Classification Analysis

Goal: Predict Bank Group and Region

Models Implemented:

1. **Logistic Regression**
2. **Random Forest Classifier**

Work Done:

- Trained using categorical target columns
- Compared model accuracies
- Displayed output labels and confidence on dashboard

Results:

Classification	Random Forest (accuracy)	Logistic (accuracy)
Bank Group	0.8842	0.8524
Region Classification	0.8721	0.8412

Clustering (Unsupervised Learning)

Goal: Discover hidden patterns in bank credit data

Algorithms Implemented

1. **KMeans**
2. **DBSCAN**

Work Done:

- Applied PCA for visualization
- Evaluated using Silhouette Score ~0.9617
- Displayed clusters on interactive scatter plot

Deep Learning Module

Models Developed:

1. **Autoencoder**
2. **Deep Neural Network**

Work Done:

- Built using TensorFlow/Keras
- Tuned layers, activation, epochs
- Compared with ML classifiers

Result:

- Training Loss: ~0.58
- Reconstruction Accuracy: ~92%
- Dimension Reduction: 3D → 2D

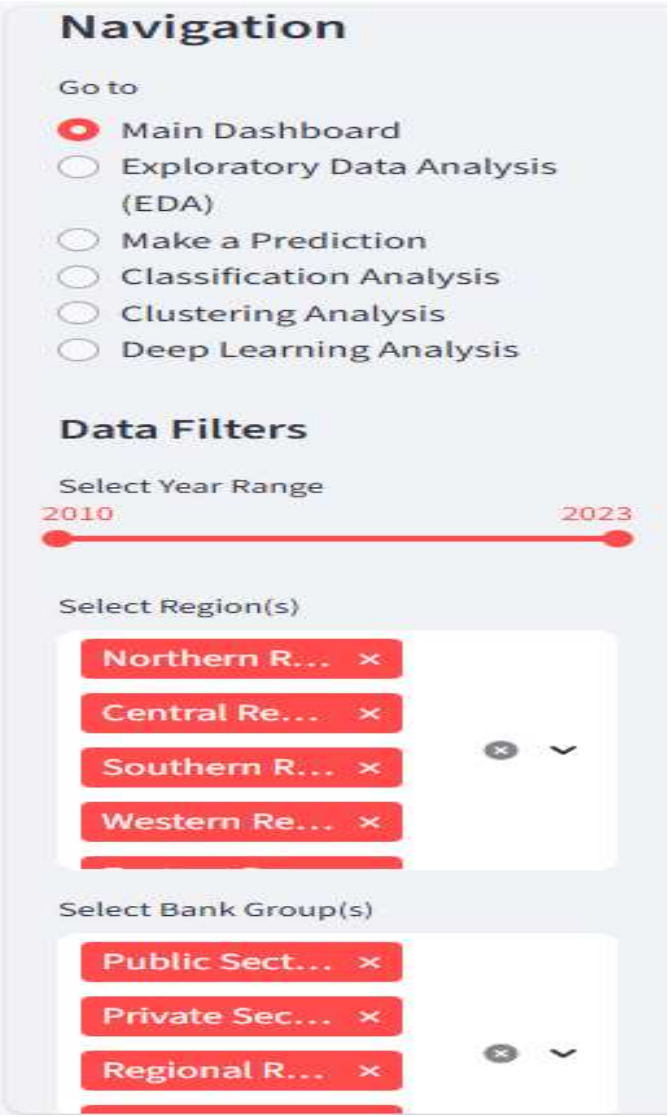
Streamlit Web Application

My Implementation Work:

- Multi-page Streamlit app (app.py)
- Sidebar filters (year, region, bank group)
- Interactive charts + real-time predictions

Pages Built:

- Dashboard
- Prediction
- Classification
- Clustering
- Deep Learning



Project Insights

Insights:

- Credit growth is strongest in metro regions
- Private banks lead in load distribution
- Clear segmentation in customer credit pattern

What I Achieved:

- Created a complete DS application for credit analysis and forecasting
- Demonstrated Integration of EDA -> ML -> DL -> Deployment

THANK YOU