SLIDE DECK

PRESENTATION

Submitted by :- Preet Kodape
email :- preetkodape9430@gmail.com
number :- 9325447962

ABOUT ME
PREET KODAPE

I am Recent Graduate Student from IIT Bombay Batch on 25 from Department of Aerospace Engineering

I have previously interned as a Marketing Intern at Dison and as a Product Management Intern at SwitchON Foundations. Now, I'm making the shift from a non-technical background to a technical role, and I'm eager to gain hands-on experience in AI/ML at companies rather than focusing solely on personal projects. I'm also passionate about badminton, drawing, and drama

KEY PARTS OF **PROJECT**

1 Creating a Synthetic Dataset to Simulate Realistic Credit Behavior

2 Training and Evaluating the Multi-Class Classification Model

3 Bonus
Proposing Interventions and Enhancing with Streamlit

Main Objective

Simulate a dataset and build a machine learning model to predict whether a customer's credit score will increase, decrease, or remain stable in the next 3 months

LOGIC BEHIND EACH **COLUMNS (PART 1)**

Assumtions:-

Customer Data: customer_id: A unique identifier for each customer, ranging from 1 to 25,000.
Gender: The dataset includes both Male and Female customers.
Location: The customers are from the top 100 cities in India, ensuring a diverse geographical representation.

Age Range:

- Customers' ages are divided into the following ranges:
  - 21-30: 55% of customers
  - 31-40: 18% of customers
  - 41-50: 12% of customers
  - 51-60: 8% of customers
  - 61-70: 7% of customers

Monthly Income: connected to Age Group

- Age and Income Range:
  - 21-30 years: ₹20k - ₹50k
  - 31-40 years: ₹50k - ₹100k
  - 41-50 years: ₹100k - ₹150k
  - 51-60 years: ₹150k - ₹200k
  - 61-70 years: ₹200k+

Current Outstanding Calculation

- Tenure and EMI Outflow:
  - Short-term loans (1-12m) common in younger age groups.
  - $Z = T \times EMI + EMI \times (0.08 - 0.1)$: T = Tenure
- Tenure Probability: Short term loans
  - 21-30 years: 70% (1-12 months).
  - 31-40 years: 60%
  - 41-50 years: 50%
  - 51-60 years: 40%
  - 61-70 years: 70%

Monthly EMI Outflow :-EMI as a % of Monthly Income. 30% of the individuals in each age group follow the 30% rule for EMI outflow.

- 21-30 years: 20%
- 31-40 years: 30%
- 41-50 years: 40%
- 51-60 years: 50%
- 61-70 years: 60%

Credit Utilization Ratio

- Credit Usage: assumes that individuals follow the 30% credit usage rule based on age.
- Credit Utilization Ratio= usage / Limit
- Utilization Based on Age: follow 30% rule
  - 21-30 years: 45%
  - 31-40 years: 55%
  - 41-50 years: 60%
  - 51-60 years: 70%
  - 61-70 years: 80%

Number of Open loans

- 7 types of main loans :- Personal, Home, Auto, Education, consumer , Gold, Credit card
- Age wise loan max
  - 21-30 :- 3
  - 31-40 :- 4
  - 41-50 :- 3
  - 51-60 :- 2
  - 61-70 :- 1

Total Credit Limit

- **General Assumption:** The total credit limit is typically 2-3 times the monthly income.
- **Example:** For a salary of ₹50,000, the credit limit would be between ₹1,00,000 and ₹1,50,000.
- **Rounding:** The limit is rounded to the nearest multiple of 5 or 10 for simplicity (e.g., ₹1,67,038 rounded to ₹1,70,000).
- **Code Logic:** The function generates the credit limit by multiplying the income by a random value between 2 and 3, then rounding the result to the nearest 10,000

Repayment History Score (0–100) by Age Group

Age Group	Excellent Score (85-100)	Good Score (70-84)	Fair Score (50-69)	Poor Score (0-49)	Average Score
21-30	25%	35%	30%	10%	72
31-40	35%	40%	20%	5%	78
41-50	45%	35%	15%	5%	82
51-60	55%	30%	12%	3%	86
61-70	65%	25%	8%	2%	89

Months Since Last Default by Age Group

Age Group	Never Defaulted	Recent Default (0-6 months)	Recovery Period (6-24 months)	Strong Recovery (24+ months)	Average Months Since Default
21-30	70%	12%	10%	8%	18 months
31-40	75%	8%	9%	8%	22 months
41-50	82%	5%	7%	6%	28 months
51-60	88%	3%	5%	4%	35 months
61-70	92%	2%	3%	3%	42 months

DPD (Days Past Due) by Age Group

Age Group	% with DPD 000 (All On Time)	% with DPD 015-030 (Minor Delays)	% with DPD 060+ (Major Delays)	Key Influencing Factors
21-30	65%	25%	10%	Young, low income, 3 loans, only 20% follow EMI rule, lowest repayment scores, highest recent defaults
31-40	75%	18%	7%	Moderate income, 4 loans, 30% follow EMI rule, better repayment, fewer recent defaults
41-50	83%	13%	4%	Higher income, 3 loans, 40% follow EMI rule, higher repayment scores, fewer defaults
51-60	90%	8%	2%	High income, 2 loans, 50% follow EMI rule, strong repayment, very few defaults
61-70	95%	4%	1%	Highest income, 1 loan, 60% follow EMI rule, best repayment scores, almost no defaults

Number of Hard Inquiries Last 6 Months by Age Group

Age Group	Excellent Repayment (85-100)	Good (70-84)	Fair (50-69)	Poor (0-49)	Notes
21-30	0-1	1-2	2-4	4-8	Youngest, most credit-seeking, highest stress if poor score
31-40	0-1	1-2	2-3	3-6	High credit activity, but more stability than 21-30
41-50	0	1	1-2	2-4	Credit needs stabilize, fewer inquiries overall
51-60	0	0-1	1-2	2-3	Rarely seek new credit, even with fair/poor scores
61-70	0	0-1	1	1-2	Very low credit-seeking across all scores

Recent Credit Card Usage (Last 3 Months) by Age Group

Age Group	Excellent Repayment (85-100)	Good (70-84)	Fair (50-69)	Poor (0-49)	Notes
21-30	10-25%	20-35%	30-50%	50-80%	Young, impulsive, often high utilization if stressed
31-40	8-20%	15-30%	25-45%	40-75%	Still high spenders, but more stable than 21-30
41-50	5-15%	10-25%	20-40%	35-65%	Usage moderates as income and discipline rise
51-60	3-12%	8-20%	15-30%	25-50%	Most conservative, lowest utilization overall
61-70	2-10%	5-15%	10-25%	15-35%	Very low usage, even with lower scores

Recent Loan Disbursed Amount by Age Group and Repayment Score

Age Group	Income Range (₹/month)	Excellent Repayment (85-100)	Good (70-84)	Fair (50-69)	Poor (0-49)	Notes
21-30	20k–50k	2–5 lakh	1–3 lakh	0.5–2 lakh	0–0.5 lakh	Young, low income, less history, more rejections if poor score
31-40	50k–100k	5–12 lakh	2–7 lakh	1–4 lakh	0–1 lakh	More income, higher approvals, but defaults still penalized
41-50	100k–150k	10–18 lakh	5–12 lakh	2–6 lakh	0–2 lakh	Peak earning, best eligibility, but poor scores still limit
51-60	150k–200k	8–15 lakh	4–10 lakh	1–5 lakh	0–1 lakh	Approvals high for good scores, but age may limit tenure
61-70	200k+	5–10 lakh	2–6 lakh	0.5–3 lakh	0–0.5 lakh	High income but lower demand, lenders cautious with poor scores

TARGET CREDIT SCORE MOVEMENT **(INCREASE, DECREASE, STABLE)**

Logic Summary:

- Indicators Influencing Movement:
  - **Negative Indicators (Risk Factors)**:
    - High DPD (Days Past Due), Recent Default (within 3 months), High Credit Utilization, Multiple Hard Inquiries, Low Repayment Score, High EMI-to-Income Ratio, High Recent Loan Disbursed Amount, High Recent Credit Card Usage.
  - **Positive Indicators (Healthy Financial Behavior):**
    - On-Time Payments (DPD 0), Long Recovery Period (more than 24 months since last default), Low Credit Utilization, Few or No Hard Inquiries, High Repayment Score, Low EMI-to-Income Ratio, Low Recent Loan Disbursed Amount, Low Credit Card Usage.

TARGET CREDIT SCORE MOVEMENT **(INCREASE, DECREASE, STABLE)**

Outcome Logic:

- Decrease: When there are 4 or more negative indicators (e.g., high DPD, high credit utilization, multiple inquiries).
- Increase: When there are 4 or more positive indicators (e.g., on-time payments, low credit utilization, low EMI ratio).
- Stable: When neither "Decrease" nor "Increase" criteria are met, indicating stable behavior.

Code Logic:

The target credit score movement is determined based on the balance of positive and negative financial indicators.

MODEL TRAINING AND EVALUTATION
PART 2

Data Preprocessing

- Synthetic Dataset: 25,000 rows with customer credit behavior.
- Key Features: monthly_income, credit_utilization_ratio, dpd_last_3_months, repayment_history_score, etc.
- SMOTE applied for class imbalance.

Models Used : RandomForest, XGBoost, LogisticRegression, LightGBM.

Hyperparameter Tuning : GridSearchCV to find optimal parameters for each model.

Model Performance

Model	Accuracy	F1-Score
LightGBM	96.10%	96.17%
XGBoost	95.10%	95.20%
RandomForest	92.60%	92.85%
LogisticRegression	66.30%	69.99%

KEY INSIGHTS&
INTERVENTIONS

SHAP Analysis: Key Features

- **Decrease in Credit Score:**
  - Key Features: credit_utilization_ratio, dpd_last_3_months, num_hard_inquiries_last_6m.
- **Increase in Credit Score:**
  - Key Features: num_hard_inquiries_last_6m, credit_utilization_ratio, dpd_last_3_months.

Business Takeaways

- **Risk Management:** Identifying high-risk customers enables proactive mitigation to reduce losses.
- **Customer Segmentation:** Tailored offerings based on age, income, and behavior improve satisfaction and retention.
- **Predictive Insights:** The model optimizes credit decisions, such as loan approvals and credit limits.
- **Scalability:** The model can be adapted to real-world data for AI-driven decision intelligence.

BONUS PART

PRODUCT/ POLICY INTERVENTIONS

1 High-Risk Segments

- **Financial Counseling:** Help manage credit and reduce DPD.
- **Credit Limit Reduction**: Lower limits to reduce stress.
- **Prevention of New Credit:** Stop new loans/credit until improvement.
- **Payment Restructuring:** Offer easier repayment options.
- **Targeted Marketing**: Provide low-interest loans and balance transfers.

2 High-Opportunity Segments

- **Credit Limit Increase**: Reward responsible behavior with higher limits.
- **Pre-Approved Loans**: Offer loans with attractive terms.
- **Rewards Programs**: Cash-back or points for good behavior.
- **Loan Refinancing:** Offer lower rates for good payers.
- **Lower Interest Products**: Provide lower rates to improve credit scores.

THANK YOU

Github

Linkedin