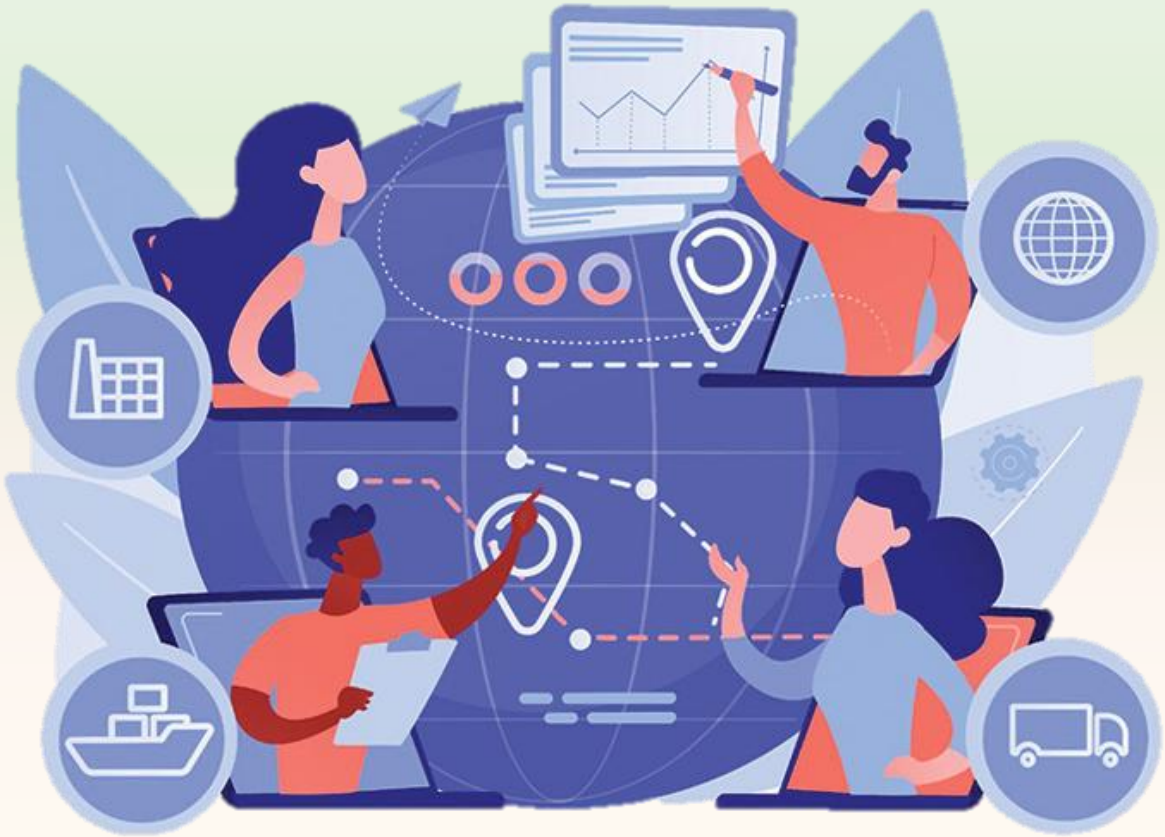
Knowledge Graph Integrated Credit Risk Assessment

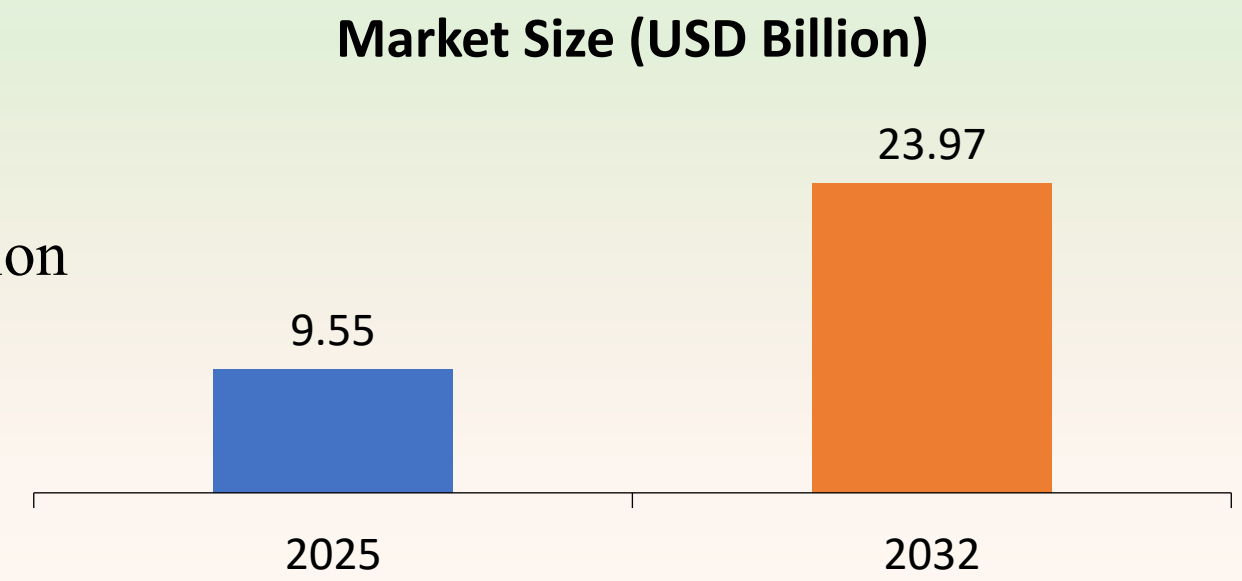
Prof. Manoj Kumar Tiwari
Director, IIM Mumbai
Professor (HAG) IIT Kharagpur (on lien)



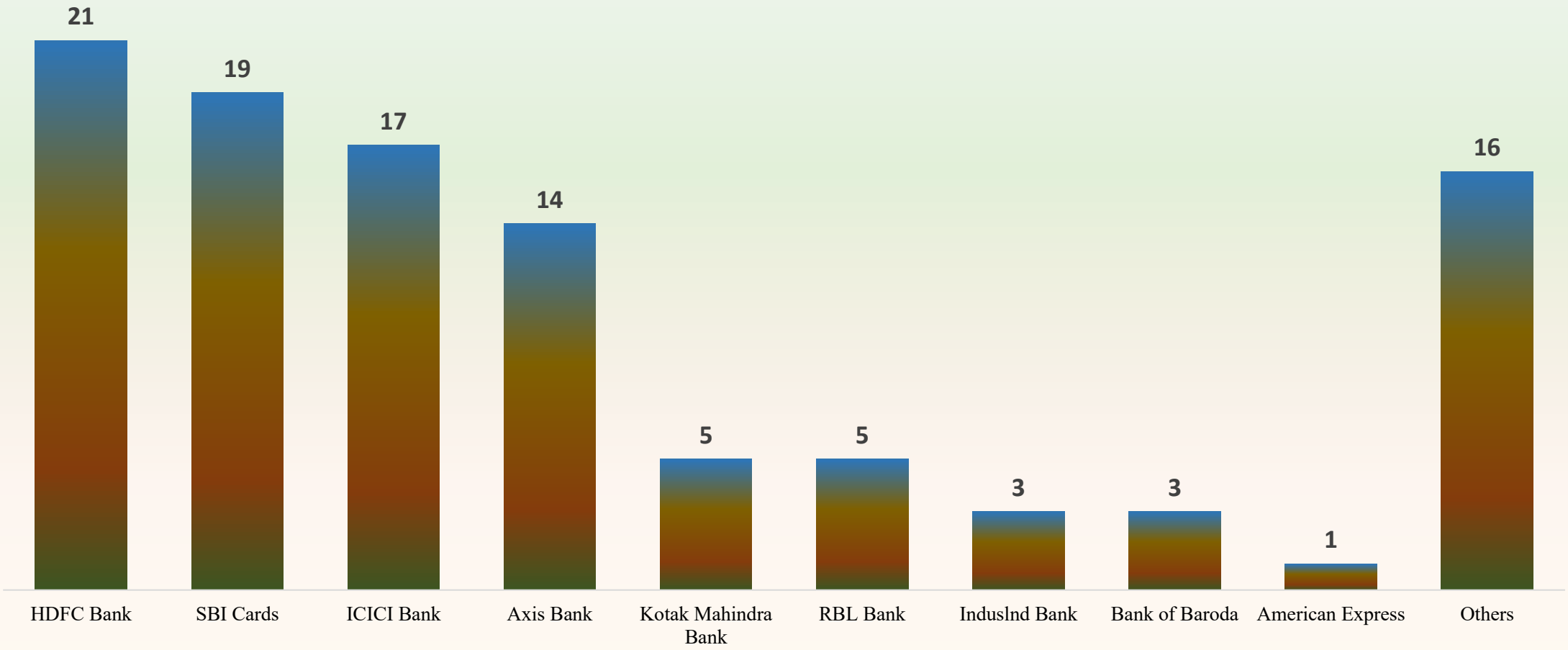
Market Size & Growth For Credit Risk Assessment

The credit risk assessment:

- Market Value in 2025: USD 9.55 billion
- Projected to reach by 2032: USD 23.97 billion
- CAGR: 14.1%
- Drivers: AI, ML, Big Data, Regulatory Mandates



Credit Card Market Share by Bank (in %) in India 2024



Market share of credit cards in India as of October 2024, by bank (in %)

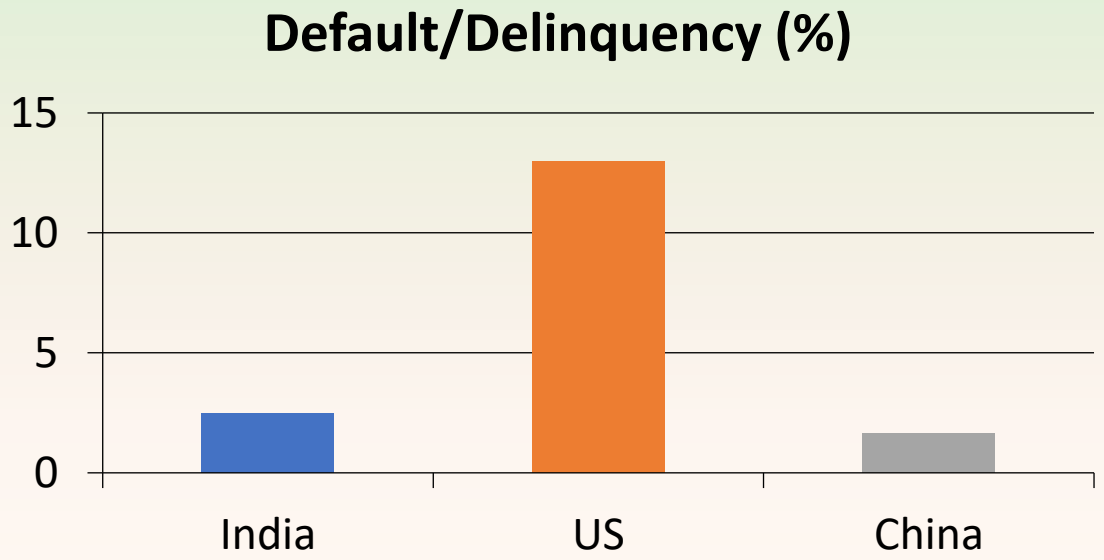
Penetration Rate of Credit Cards in India 2014-2029 (in %)

Credit card penetration measures the percentage of transactions completed using credit cards compared to other payment methods.

Penetration rate of credit cards in India from 2014 to 2029 (in %)

Default Rate Comparison – India vs US vs China

- India Corporate Default Rate: <2.5%
- US Consumer Loan Delinquency Increase: 13%
- China Bad Loan Ratio: 1.66%



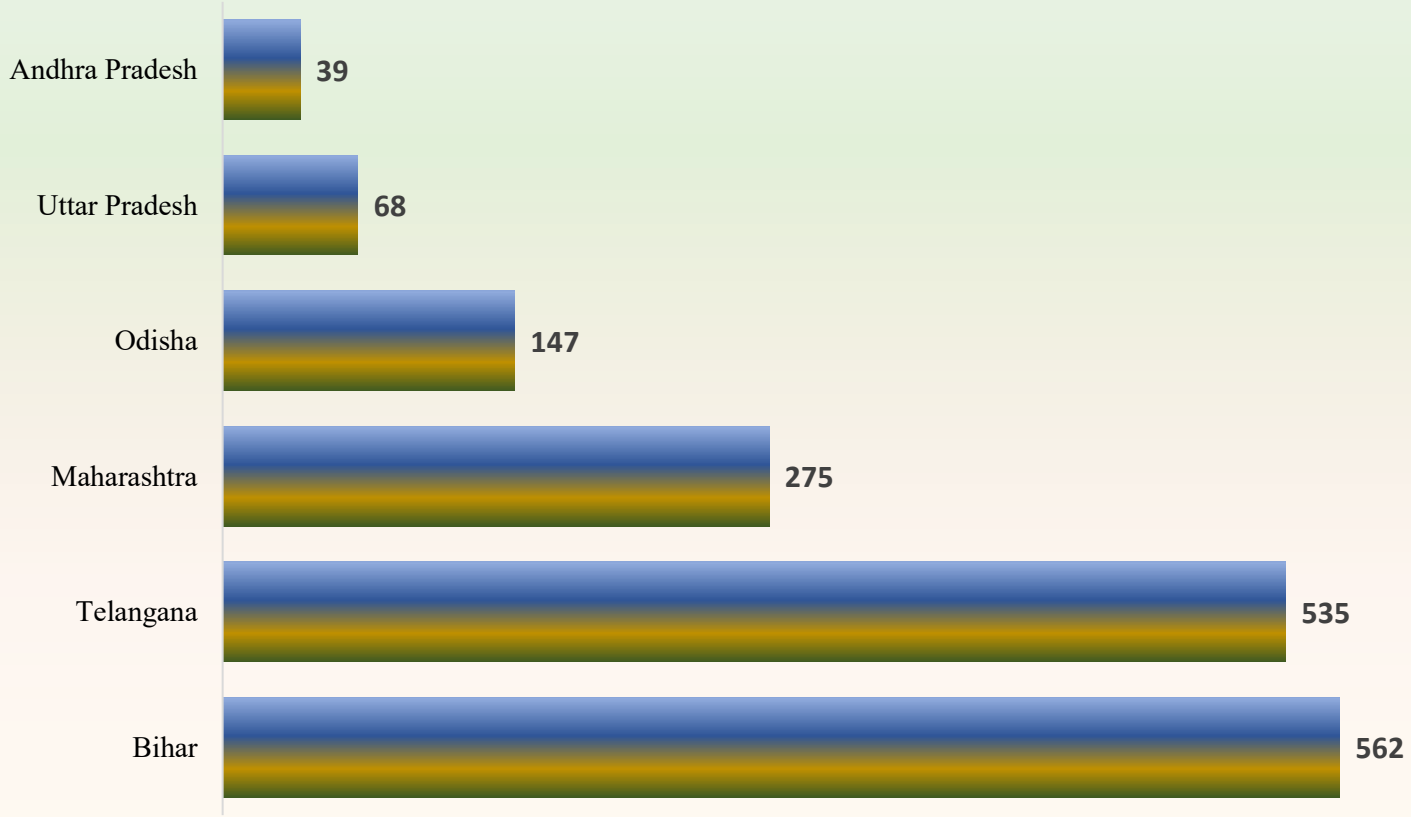
Key Challenges

- 25–28% of banks struggle with data quality issues
- 80% of customers fear data privacy breaches
- Regulatory tightening increases compliance burden
- Many lenders rely on outdated risk engines

Credit Risk Statistics

- Days Sales Outstanding (DSO) increased for 57% of businesses
- US bankruptcy filings rose 33.5% (12 months ending Sept 2024)
- 39% of invoices paid late
- 57% higher accounts receivable (AR) difficulty with >30 day terms

Credit And Debit Cards Incidents Reported In India 2022 By Leading State



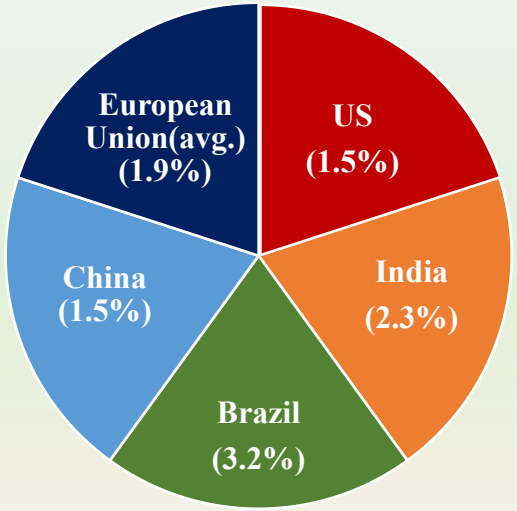
Number of credit and debit card fraud incidents reported across India in 2022, by leading state

Different Types of Financial Risks

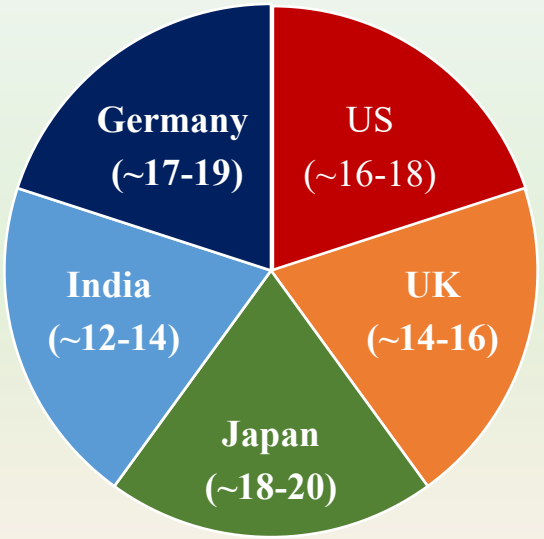
- Credit Risk
- Market Risk
- Liquidity Risk
- Operational Risk
- Interest Rate Risk
- Concentration Risk

COUNTRY-WISE RISK STATISTICS

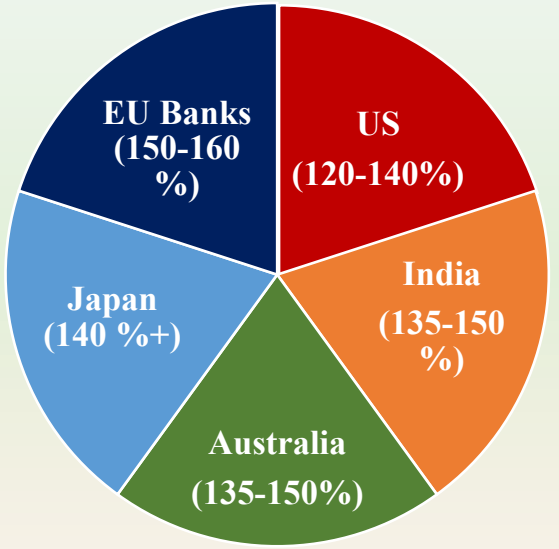
CREDIT RISK
(Non-Performing Loan Ratio, NPL%)



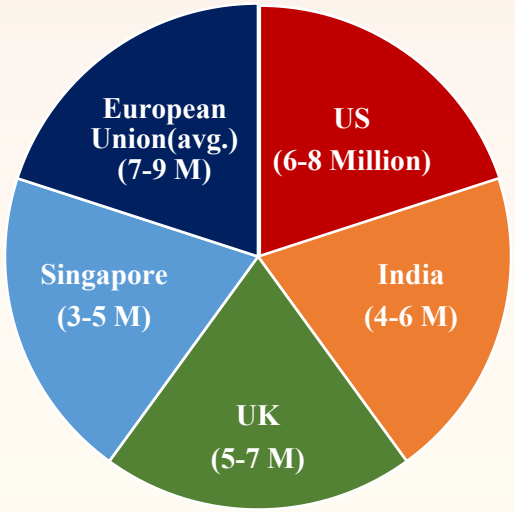
Market RISK
(VOLATILITY INDEX)



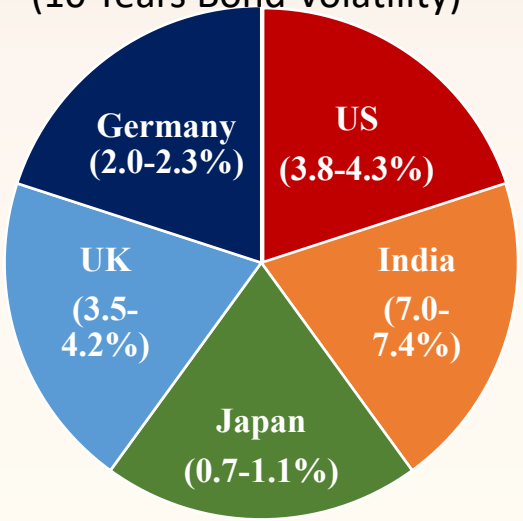
Liquidity RISK
(LCR %)



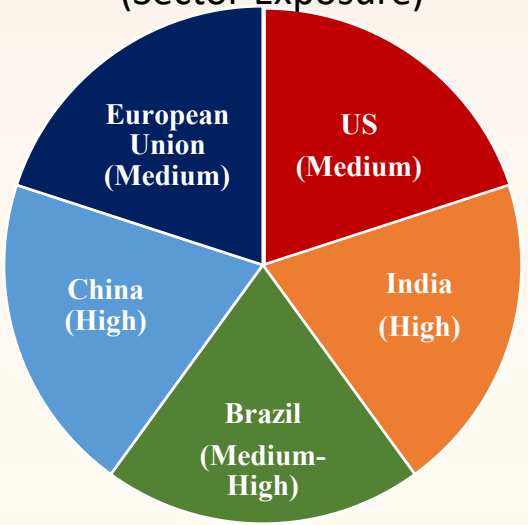
Operational Risk
(Losses Per $1B Assets)



Interest Rate Risk
(10 Years Bond Volatility)



Concentration Risk
(Sector Exposure)



Credit Risk

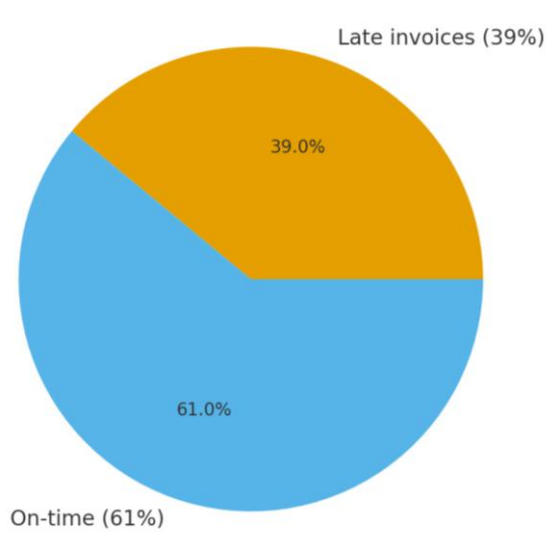
- Risk of loss due to borrower default
- Affects banks, Non-Banking Financial Companies (NBFCs), fintech
- Includes Probability of Default (PD), Loss Given Default (LGD), Exposure at Default (EAD)

Credit Risk

Global Credit Risk

- Global NPL Ratio (2024): ~6%
- SME loan default rate: 8–12%
- Credit card delinquency rising globally

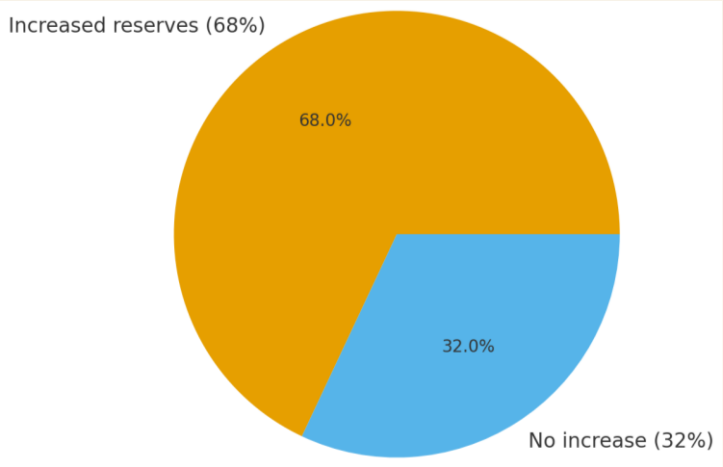
Late Payment



Credit Risk in India

- India NPA ratio: 3.9% (RBI 2024)
- MSME shows highest stress
- Surge in digital lending frauds

Bad Debt Reserve Changes



Different Types of Credit Risk

- **Default Risk**

  A borrower fails to repay principal or interest.

- **Concentration Risk**
  - Excessive exposure to a single borrower, sector, or region.
  - Leads to high vulnerability during downturns.

- **Counterparty Risk**

  Risk that a trading partner or financial institution fails to meet obligations (e.g., derivatives, interbank lending).

- **Sovereign Risk**
  - Risk that a government defaults on its debt or imposes capital controls.
  - Affects foreign lending and investments.

- **Settlement Risk**

  Risk that a payment or transaction fails to settle even after the counterparty has received funds.

- **Migration (Downgrade) Risk**

  Risk of borrower's credit rating worsening, increasing expected losses.

- **Recovery Risk**

  Uncertainty about how much (if any) money can be recovered after default.

Why Credit Risk Assessment Matters?

- Prevents financial losses
- Ensures safe lending
- Supports economic stability

Modern Credit Risk Approaches

- AI/ML models
- Alternative data
- Knowledge graph scoring
- Behavioral analytics

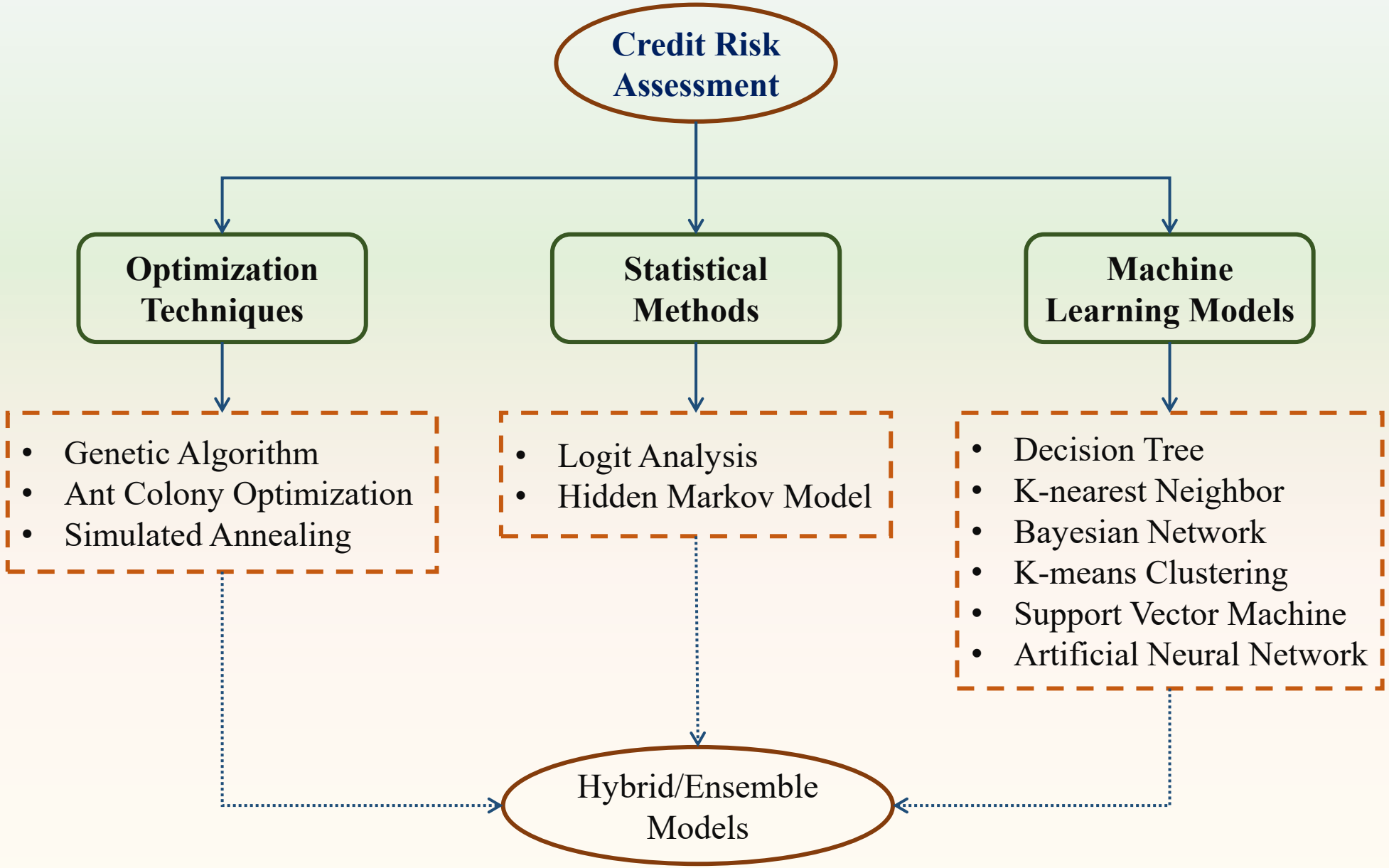
Traditional Credit Risk Assessment

- Income & collateral verification
- Bureau scores
- Manual underwriting

Credit Scoring Components

- Probability of Default (PD)
- Loss Given Default (LGD)
- Exposure at Default (EAD)
- Expected Credit Loss (ECL)

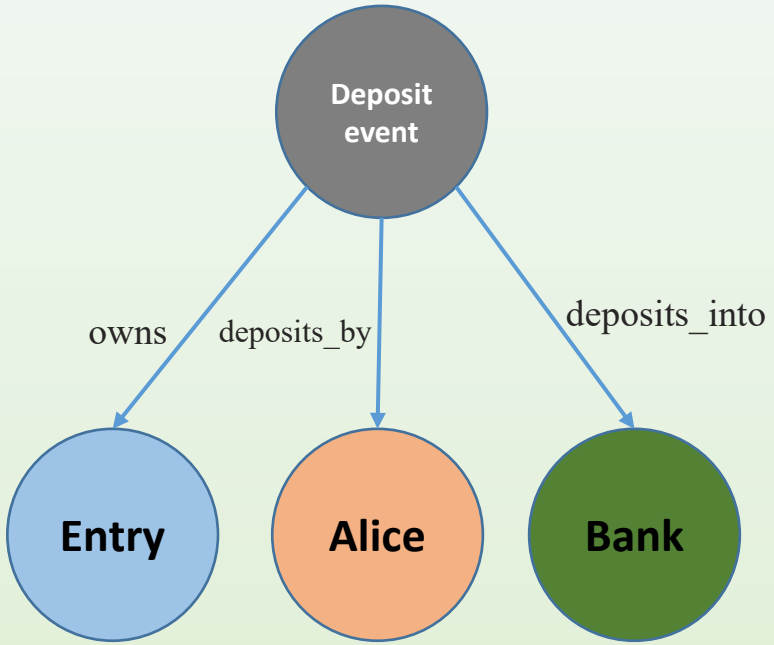
Traditional Credit Risk Assessment



Limitations of Traditional Credit Scoring

- Static scorecards
- Limited relational intelligence
- Poor for new-to-credit customers
- Vulnerable to fraud networks

Knowledge Graph



- A type of database that represents knowledge in a structured way by organizing and linking data points through relationships.
- Used to connect different pieces of information, allowing for more efficient and effective data processing and analysis

1 **Nodes & Relationships**
Nodes represent entities while relationships define connections between entities.

2 **Semantic Web**
Knowledge graphs leverage linked data principles to establish semantic connections.

3 **Knowledge Representation**
Structured data in RDF (Resource Description Framework) format enables efficient knowledge representation.

Benefits of Using a Knowledge Graph

- **Improved Data Integration:** bring data from multiple sources together into a single unified structure, making it more accessible and easier to work with.
- **Flexible Data Connectivity:** Easily connect to and integrate data from a variety of structured and unstructured sources
- **Contextual Understanding:** Discovering hidden insights and patterns that are not apparent in traditional databases.
- **Efficient Queries:** Perform complex queries efficiently and retrieve relevant results with ease

Knowledge Graph (KG)

Components of a Knowledge Graph

Entities

Represents real-world things such as people, places, concepts, or objects.

Attributes & Properties

Characteristics and qualities that describe entities and define their relationships.

Ontologies

Formal description of concepts and their relationships within a specific domain.

Graph Database

A storage system that enables efficient traversal and querying of knowledge graph data.

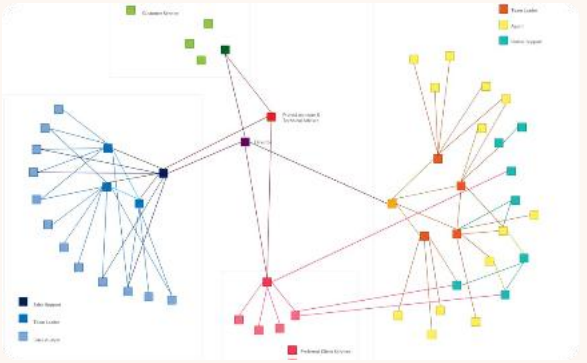
Creating and Organizing a Knowledge Graph

Building Blocks

- Identifying entities
- Defining relationships
- Establishing semantic connections

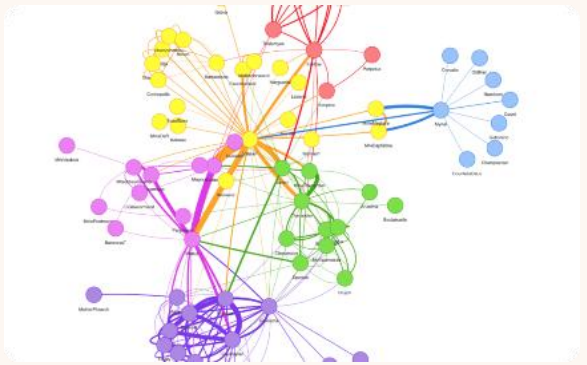
Data Integration

Aggregate and harmonize data from multiple sources

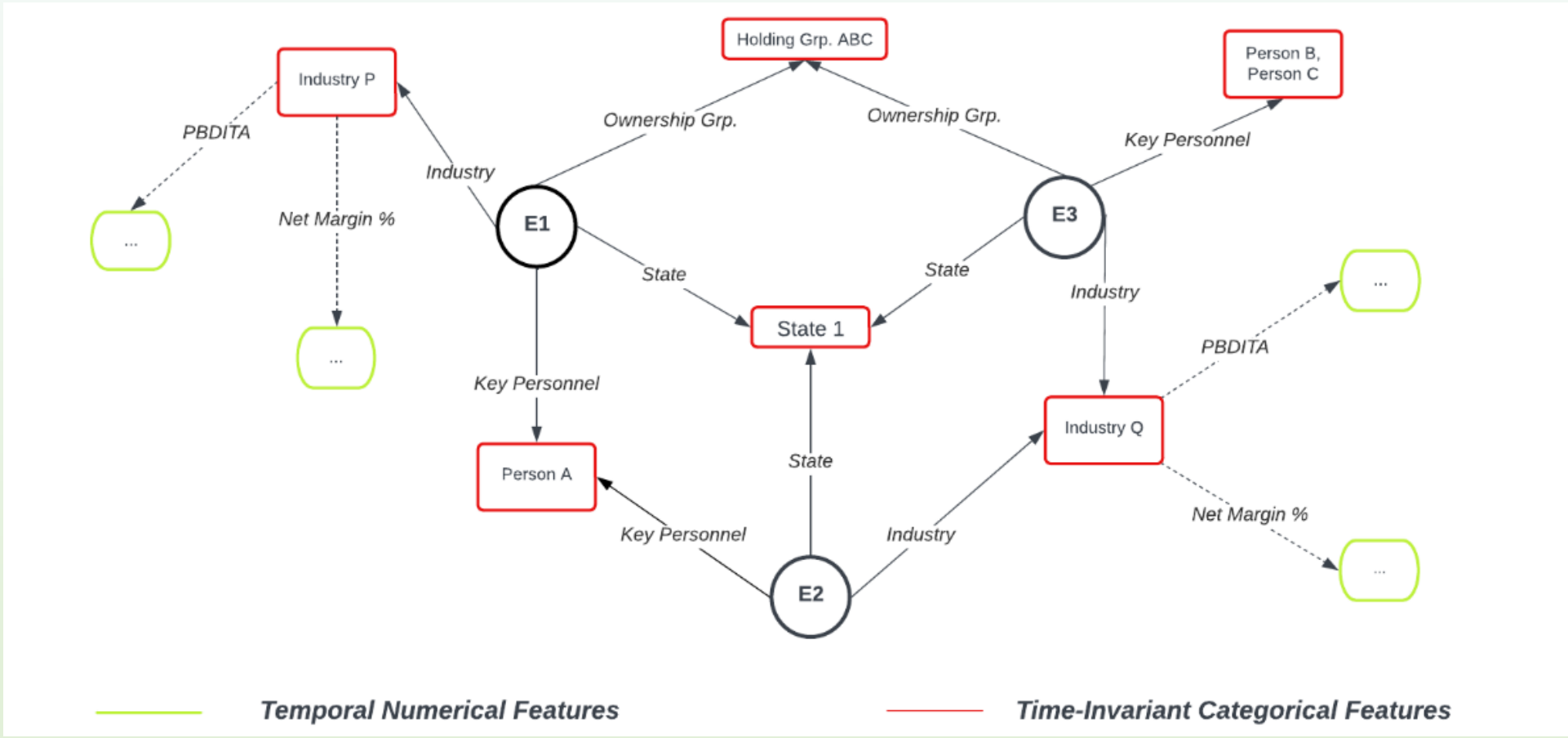


Visualization

Use modern graph visualization tools (plotly) for better comprehension



Examples of Knowledge Graph



Knowledge graph using the CMIE database for MSMEs

CMIE: Centre for Monitoring Indian Economy Pvt. Ltd.
E1, E2, and E3 are the enterprises
PBDIT: Profit Before Depreciation, Interest, and Tax

Why Use KGs for Credit Risk?

- Detect hidden relationships
- Fraud ring detection
- Real-time risk scoring
- Graph ML models capture network effects

Graph of entities connected by relationships

- Nodes: Borrowers, companies, accounts, devices
- Edges: Money flow, shared address, mutual directors

KG Integrated Credit Risk Assessment: Real-World Applications

JPMorgan Chase – Graph-based Counterparty & Credit Risk

- JPMorgan built an internal **Graph-Based Risk Engine** using entity relationships across:
  - Legal Entities
  - Subsidiaries
  - Transactions
  - Market Exposures
- Used to identify **hidden counterparty risks**, detect **cascading defaults**, and score borrower interconnectedness.
- Mentioned in their Global Technology Strategy and multiple patents.

HSBC – Knowledge Graph for Financial Crime & Credit Profiling

- HSBC uses a **Knowledge Graph + ML** system to assess:
  - Creditworthiness of SMEs
  - Fraud Detection
  - KYB (Know Your Business) Validation
- Their KG platform unifies:
  - Corporate Registries
  - Supplier Relationships
  - Trade Finance Documents
  - Payment Networks
- Used to reduce Non-Performing Asset (NPA) probabilities in trade credit.

KG Integrated Credit Risk Assessment: Real-World Applications

Alibaba / Ant Financial – SME Credit Scoring via Behavior Knowledge Graphs

- Uses a massive **behavior-behavior & entity-entity knowledge graph** to score millions of SMEs.
- Data sources:
  - Online Transaction Graph
  - Supplier Graph
  - Customer Dependency Graph
  - Logistics Graph
- 90% Of Loans Are Auto-approved Using KG + ML
- Default Prediction Accuracy Increased 30–40%

PayPal – Graph-Based Merchant Risk & Fraud Intelligence

- PayPal uses a **relationship-centric Knowledge Graph** to score merchant creditworthiness and detect high-risk financial behavior.
- KG integrates:
  - Merchant Identities
  - Customer Transaction Networks
  - Device/Browser Fingerprints
  - Chargeback Histories
  - Dispute/Complaint Networks
- The system identifies **synthetic merchants, high-risk clusters, and fraud rings** while improving **credit decisioning for PayPal Working Capital loans**.
- PayPal reports that KG-driven behavioral patterns significantly reduce **default rates and loss exposure** in SME lending.

Fintech Startups Using KG for Credit Risk Assessment

Crediwatch (India) – KG-Based Business Credit Reports

- Builds business credit intelligence and "Trust Score" ratings for MSMEs and corporates, using alternative and public data sources
- Uses knowledge graphs to connect:
  - ROC Filings
  - Court Cases
  - GST Data
  - News Sentiment
  - Director Networks
- Used by ICICI, HDFC, Axis Bank, and Non-Banking Financial Companies (NBFCs) in India.

martini

martini.ai (USA) – AI-native corporate credit analytics

- Uses a **proprietary Knowledge Graph** + agent-based AI to produce:
  - Real-time **credit ratings**
  - **Probability of Default (PD)** curves
  - Credit risk reports
- Use a **knowledge graph combined with real-time market monitoring**.

Fintech Startups Using KG for Credit Risk Assessment

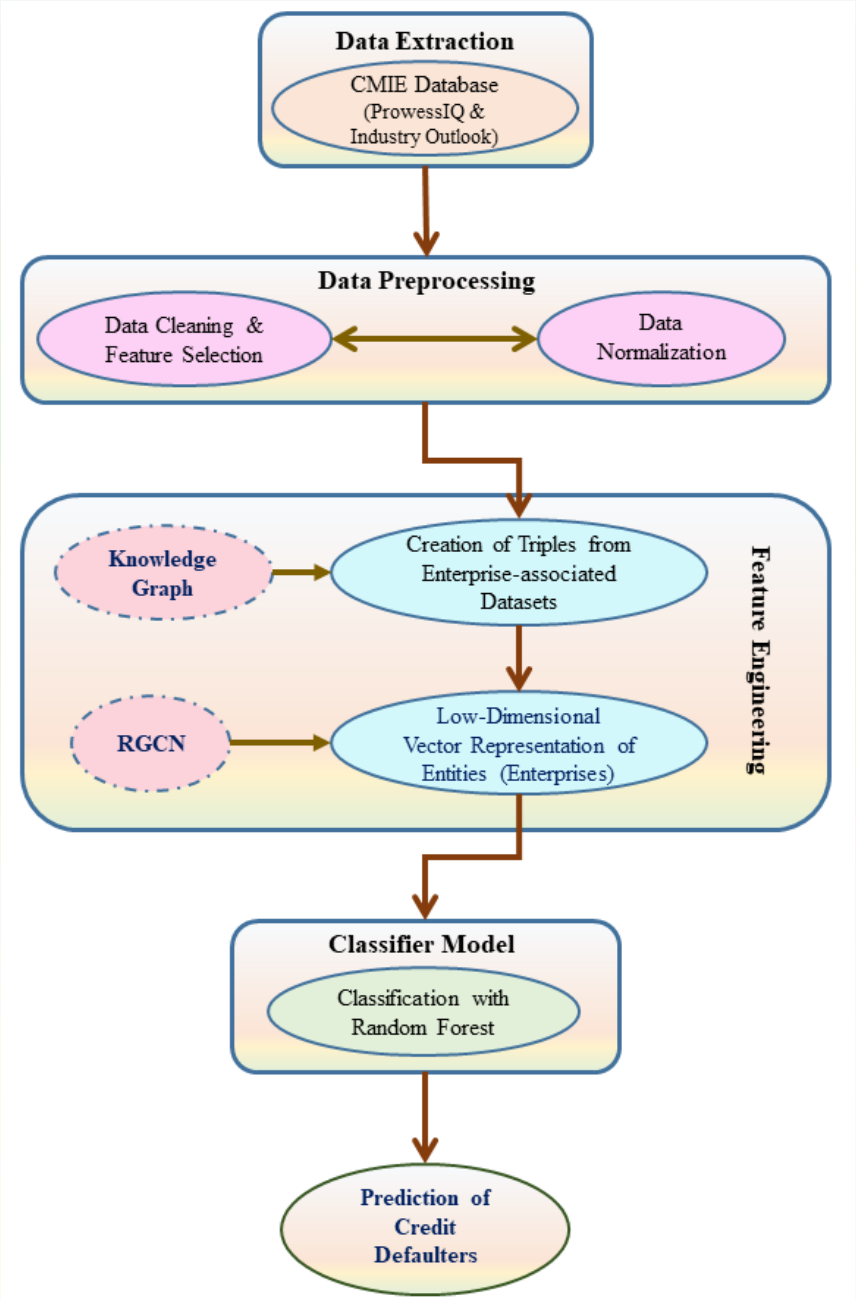
IceKredit (China/Global) – SME & individual credit scoring with graph analytics

- Uses **graph analytics** to build:
  - Credit scoring engine for **SMEs & individuals**
  - **Customer 360° Knowledge Graph** linking entities & relationships

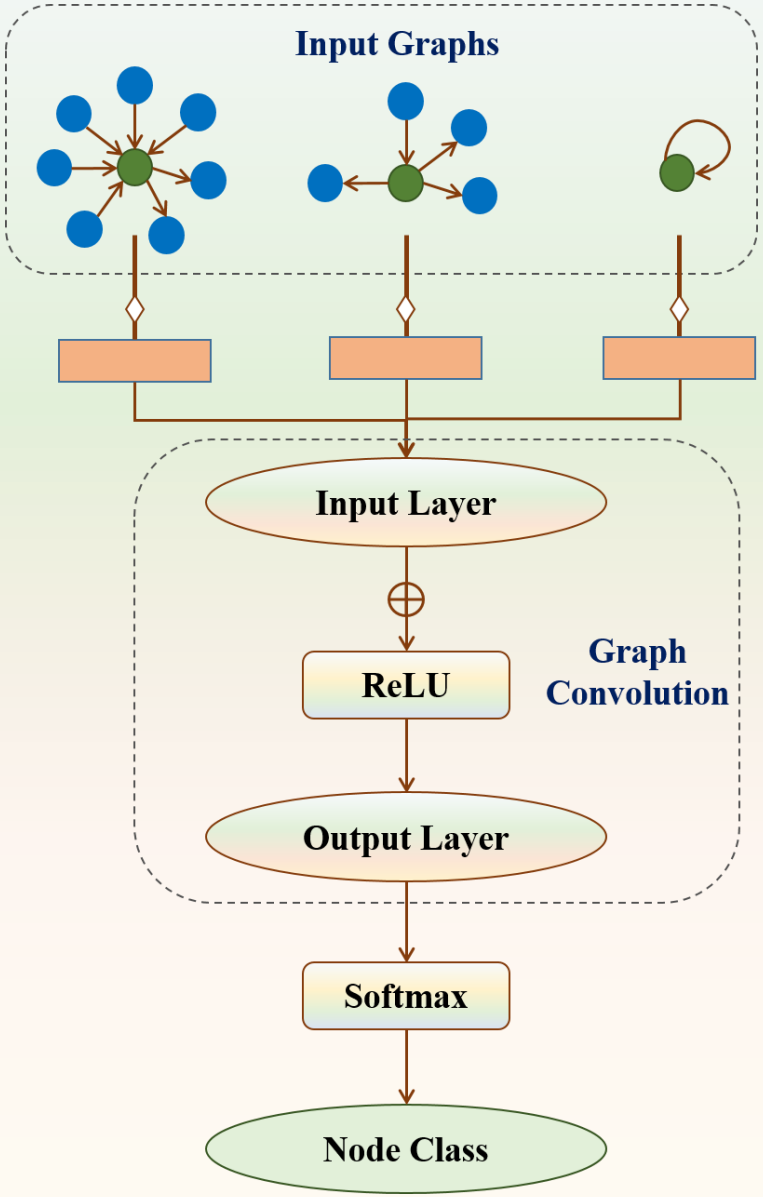
- Graph models improve:
  - Fraud detection
  - SME credit evaluation & default prediction

IceKredit

Framework for Credit Risk Assessment



Relational Graph Convolutional Network (RGCN)



RELU = Rectified Linear Unit

Experiments and Analysis

CMIE (Centre for Monitoring Indian Economy) database

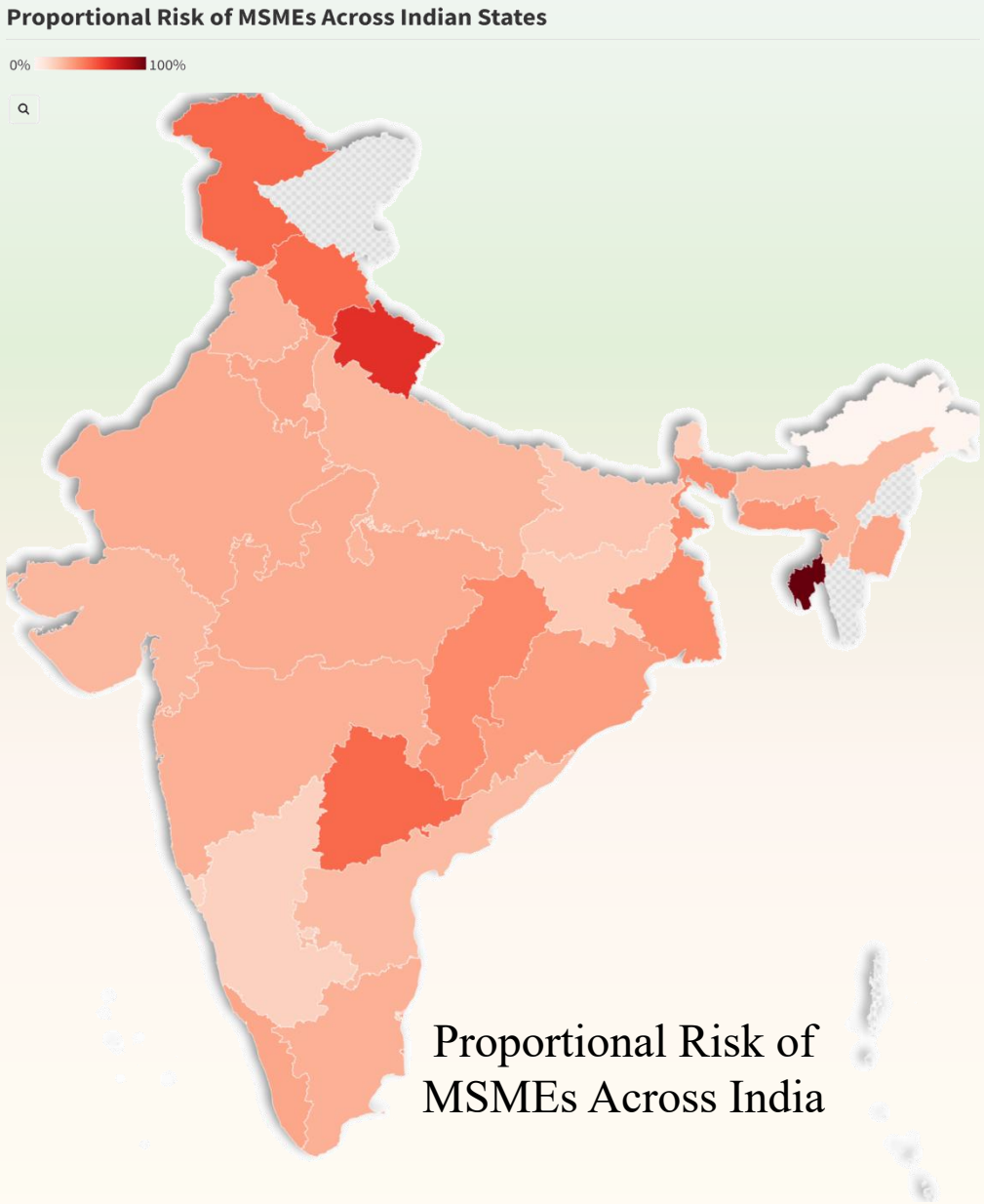
- ProwessIQ
  (Time-invariant features including financial statements, ratios, and other critical data related to Indian companies)
- Industry Outlook
  (Industry-wise temporal features, including forecasts related to the Indian economy, market trends, industry-specific analyses, PBDIT, Net Margin Percent)

- Database spanning six years (2016-2021)
- 5477 MSMEs
- 32 states and union territories
- 305 cities
- 32,862 samples
- 53 features
- Each enterprise consists
  - 10 time-invariant categorical features
  - 3 industry-wise temporal features
- 509 unique ownership groups
- 144 unique types of industries
- 555 unique NIC-group (National Informatics Centre)

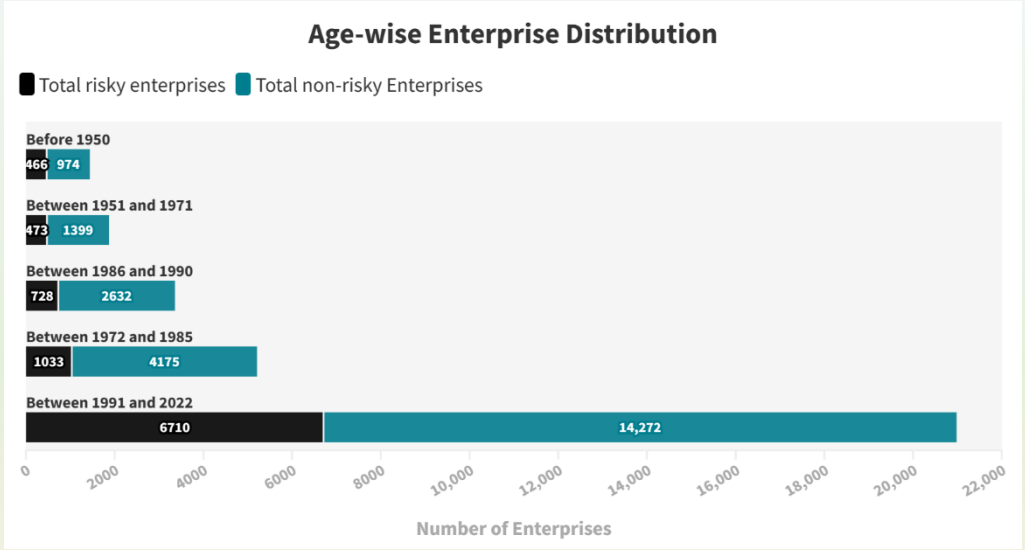
Year	Non-default	Default	Ratio (Default/Non-default)	Default Percentage
2016	3821	1656	0.43	30.23
2017	3896	1581	0.40	28.87
2018	3898	1579	0.40	28.83
2019	3869	1608	0.41	29.36
2020	3947	1530	0.38	27.93
2021	4021	1456	0.36	26.58

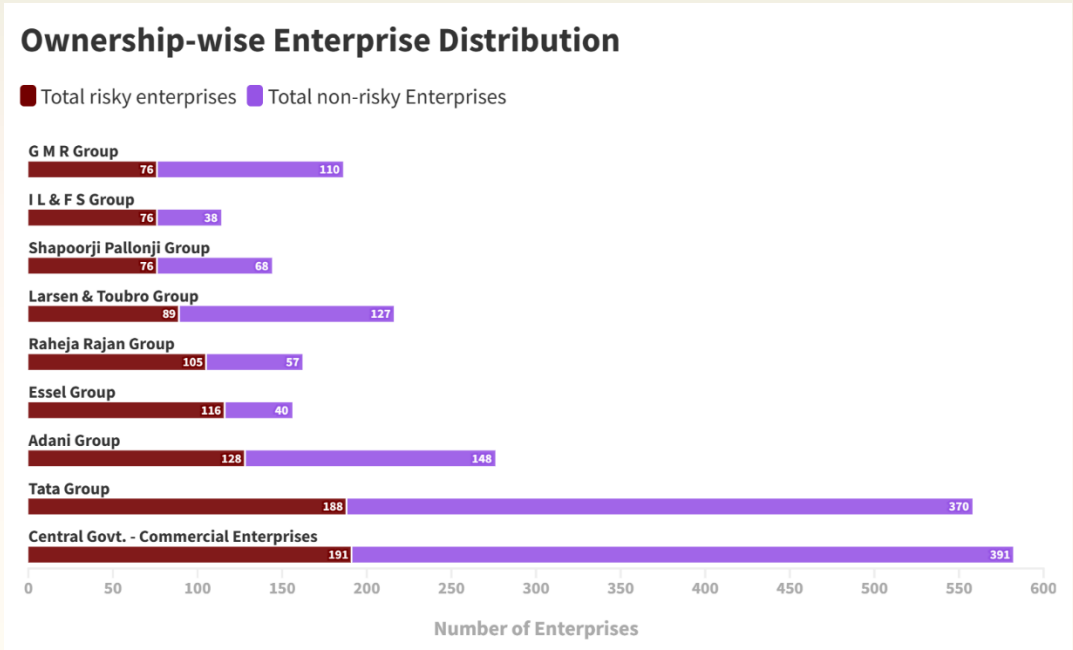
Year-wise statistics of the dataset

Explanatory Data Analysis

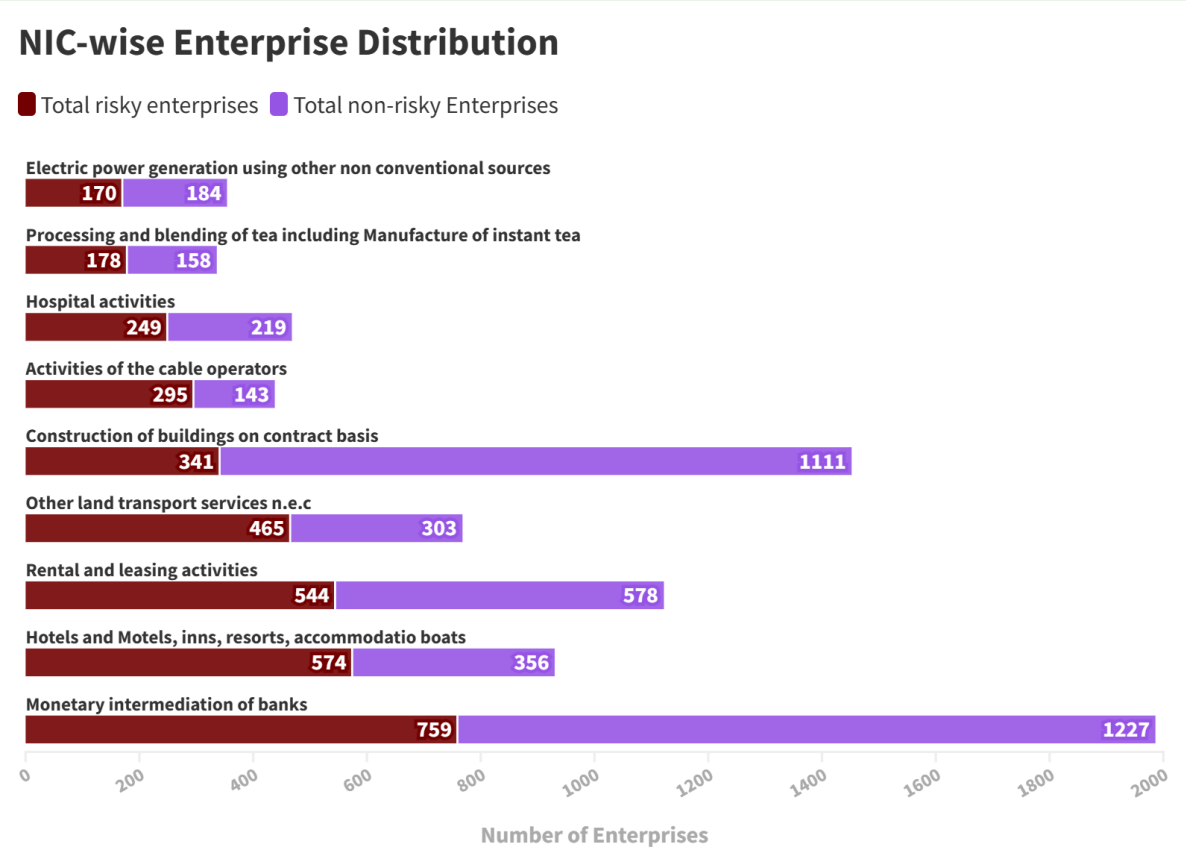


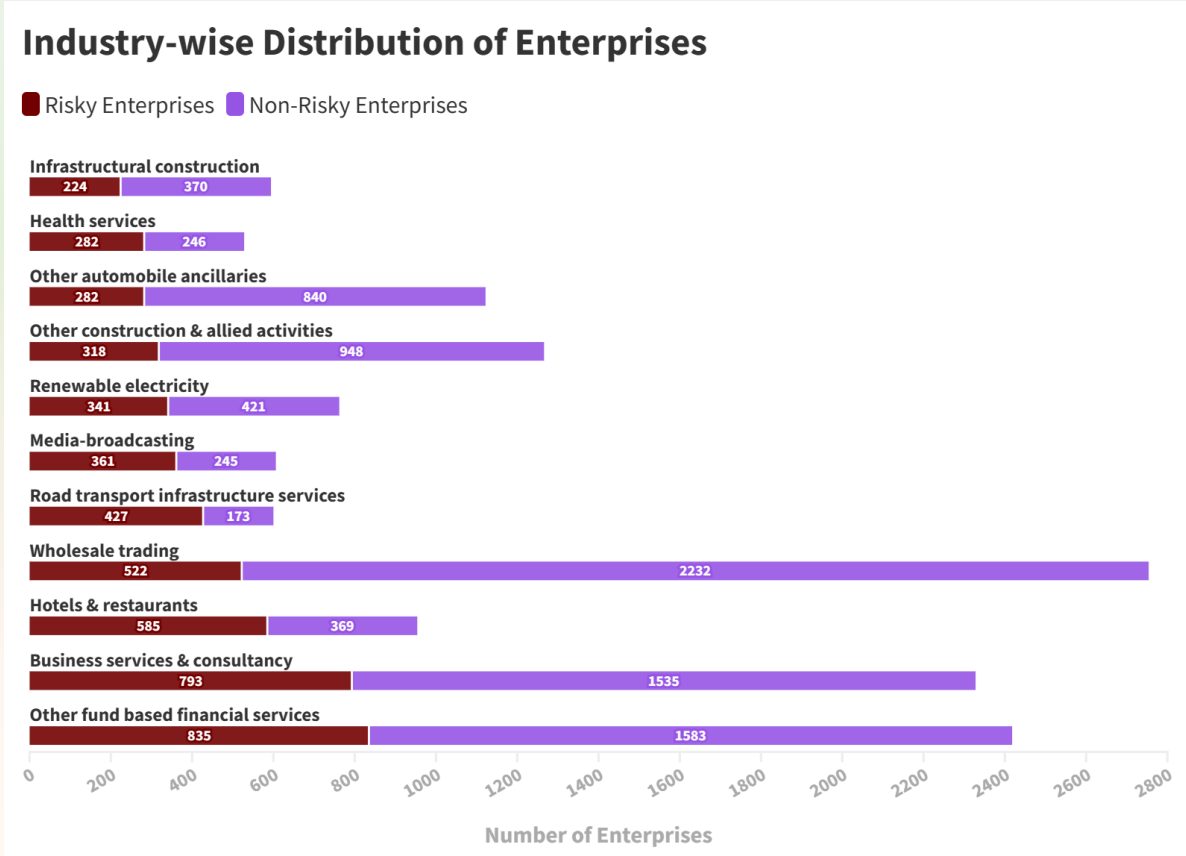
Proportional Risk of MSMEs Across India





Explanatory Data Analysis

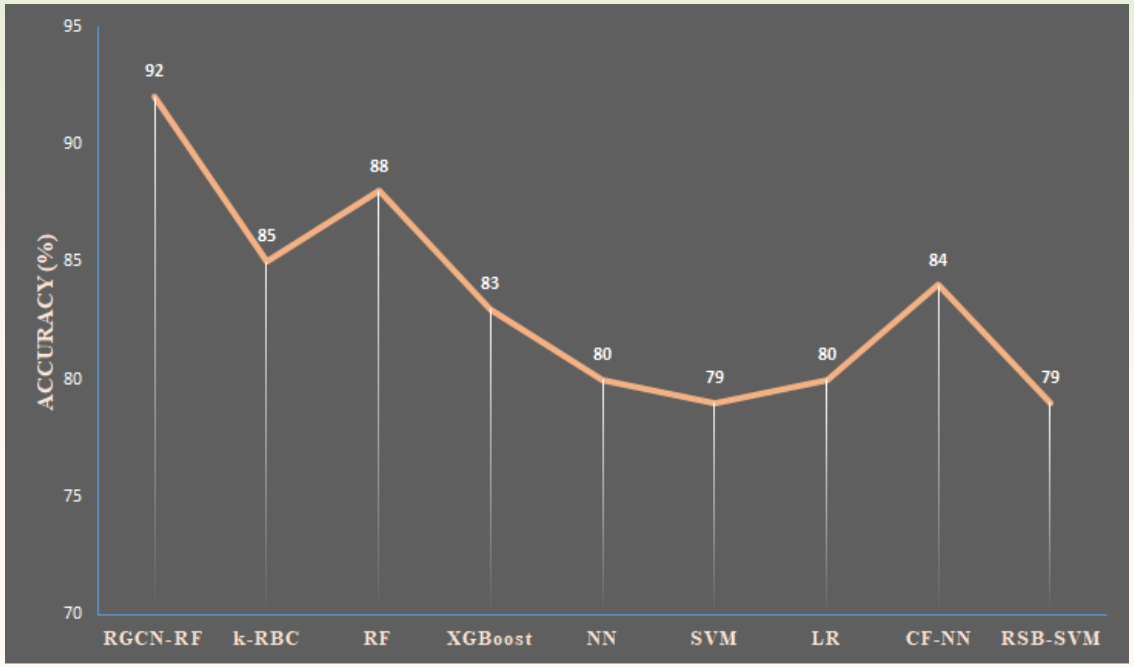




Results and Discussions

Model	Accuracy	ROC AUC	F1 Score
RGCN-RF	**0.92**	**0.92**	**0.94**
TransE-RF	0.88	0.88	0.91
One hot encoded RF	0.88	0.88	0.88
RGCN-ExtraTreeClassifier	0.87	0.87	0.89
TransE-ExtraTreeClassifier	0.87	0.87	0.89
RGCN-KNeighborsClassifier	0.85	0.85	0.88
TransE-KNeighborsClassifier	0.85	0.85	0.88
RGCN-BaggingClassifier	0.85	0.85	0.89
TransE-BaggingClassifier	0.81	0.81	0.87
RGCN-DecisionTreeClassifier	0.84	0.84	0.87
TransE-DecisionTreeClassifier	0.75	0.75	0.82
RGCN-XGBoost	0.83	0.83	0.88
TransE-XGBoost	0.81	0.81	0.87
RGCN-LightGBM	0.73	0.73	0.82
TransE-LightGBM	0.70	0.70	0.81
RGCN-BiLSTM	0.76	0.75	0.82

Performance Matrix		Actual Values	
		Default	Non-default
Predicted Values	Default	3785	236
	Non-default	243	1213



Performance of the Classification models

ROC = Receiver Operating Characteristic
AUC = Area Under the Curve

Thank You