**Credit Risk Probability Model for Alternative Data**

by

**Jerusalem Fetene**

*July, 2025*

1. **Exploratory Data Analysis (EDA)**
2. **Overview of the Data**
3. **Summary Statistics**
4. **Distribution of Numerical Features**
5. **Distribution of Categorical Features**
6. **Correlation Analysis**
7. **Identifying Missing Values**
8. **Outlier Detection**
9. **Feature Engineering**
10. **Aggregate Features**
11. **Extracted Features**
12. **Encoding Categorical Variables**
13. **Handle Missing Values**
14. **Normalizing/Standardizing Numerical Features**
15. **Proxy Target Variable Engineering**
16. **Calculate RFM Metrics**
17. **Cluster Customers using k-means**
18. **Defining and Assigning the "High-Risk" Label**
19. **Integrating the Target Variable**
20. **Model Training and Tracking**
21. **Splitting the Data**
22. **Selecting ML Models**
23. **Train the Models**
24. **Hyperparameter Tunning**
25. **Model Evaluation (accuracy, precision, recall(sensitivity), F1-score, ROC-AUC)**
26. **Unit testing**
27. **Model Deployment and Continuous Integration**